



**T.R.**  
**PAMUKKALE UNIVERSITY**  
**INSTITUTE OF EDUCATIONAL SCIENCES**  
**DEPARTMENT OF FOREIGN LANGUAGE EDUCATION**  
**ENGLISH LANGUAGE TEACHING PROGRAM**  
**MASTER OF ARTS THESIS**

**ASSESSING THE SPEAKING SKILL: AN INVESTIGATION**  
**INTO ACHIEVEMENT TESTS OF THE 9<sup>th</sup> GRADE**  
**STUDENTS IN ANATOLIAN HIGH SCHOOLS**

**Ceyda ÖZDEMİR**

**DENİZLİ – 2018**

**T.R.**  
**PAMUKKALE UNIVERSITY**  
**INSTITUTE OF EDUCATIONAL SCIENCES**  
**DEPARTMENT OF FOREIGN LANGUAGE EDUCATION**  
**ENGLISH LANGUAGE TEACHING PROGRAM**  
**MASTER OF ARTS THESIS**

**ASSESSING THE SPEAKING SKILL: AN INVESTIGATION INTO**  
**ACHIEVEMENT TESTS OF THE 9<sup>th</sup> GRADE STUDENTS IN**  
**ANATOLIAN HIGH SCHOOLS**

**Ceyda ÖZDEMİR**

**Supervisor**  
**Assoc. Prof. Dr. Turan PAKER**

## YÜKSEK LİSANS TEZ ONAY FORMU

Bu çalışma, Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Bilim Dalı'nda jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Şevki KÖMÜR (Başkan)



Üye: Doç. Dr. Recep Şahin ARSLAN



Üye: Doç. Dr. Turan PAKER (Danışman)



Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü Yönetim Kurulu'nun **12/09/2018** tarih ve **34..1..** sayılı kararı ile onaylanmıştır.



Prof. Dr. Mustafa BULUŞ

Enstitü Müdür

## ETİK BEYANNAMESİ

Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı beyan ederim.



Ceyda ÖZDEMİR

## **ACKNOWLEDGEMENTS**

I would like to express my appreciation to my thesis advisor, Assoc. Prof. Dr. Turan Paker for his never-ending support and invaluable guidance. I am also thankful to the other jury members Assoc. Prof. Dr. Recep Şahin and Prof. Dr. Şevki K m r for their helpful suggestions and comments.

I would like to thank Prof. Dr. Demet YAYLI and Asst. Prof. Dr. Selami OK for their support and comments in developing my questionnaires.

I offer sincere thanks to Denizli Directorate of National Education. They trusted and provided me with the opportunity to conduct this study in Anatolian high schools. Thanks are extended to all teachers and students for their cooperation and participation in this study.

I convey special thanks to all my friends, colleagues and students who motivated and supported me throughout this study.

Finally, I am sincerely grateful to my family for their continuous unwavering support and patience. They always believe in me and stand behind me.

## ÖZET

### **Konuşma Becerisini Değerlendirme: Anadolu Liselerinde 9. Sınıf Öğrencilerinin Başarı Sınavları Üzerine Bir Araştırma**

Ceyda ÖZDEMİR

Türkiye’de, 2014 yılından beri liselerde konuşma sınavlarının zorunlu hale gelmesiyle (Regulation for Secondary Education Institutions [Ortaöğretim Kurumları Yönetmeliği], 2014), bu sınavlar büyük çapta önem kazanmıştır. Bu çalışma Anadolu Liselerinde konuşma sınavlarının uygulama süreçlerini, 9. sınıf öğrencilerinin ve öğretmenlerinin konuşma sınavlarına karşı tutumlarını araştırmayı amaçlamaktadır. Bu araştırma 2016-2017 eğitim öğretim yılı bahar döneminde gerçekleştirilmiştir. Katılımcılar Denizli’deki altı farklı Anadolu lisesinden 358 öğrenci ve 22 öğretmendir. Araştırmanın verileri nitel ve nicel olmak üzere anket ve röportajlar yoluyla toplanmıştır. Veri analizlerinde anketler için SPSS ve röportajlar için içerik analizi kullanılmıştır. Bulgulara göre, konuşma sınavlarında standart bir prosedür yoktur. Konuşma sınavlarını öğretmenler planlayıp, bu doğrultuda öğrencileri değerlendirme ve puanlama sürecine ilişkin bilgilendirmiştir. Sınavlarda, dersin öğretmeni her bir öğrenciyi sınıf arkadaşlarının da bulunduğu sınıflarda bireysel olarak değerlendirmiştir. Bu yapılan sınavlar sonucunda hem öğretmenler hem de öğrenciler sınav süresinden memnun olduklarını ifade etmişlerdir. Ancak, puanlamada öğretmenlerin öğrencilere desteğini göstermek için fazla notlar verdiklerini belirtmiş olmaları sınavın güvenilirliğini tehdit etmektedir. Diğer bir bulgu ise, öğretmen ve öğrencilerin konuşma sınavlarına karşı olumlu tutum benimsediğidir. Öğrenciler konuşma sınavlarının İngilizcelelerini geliştirmek için faydalı olduğunu, öğretmenler de öğrencilerin seviyesini belirlemek için konuşma sınavlarına ihtiyaç duyduklarını belirtmiştir.

**Anahtar Kelimeler:** Konuşma sınavı, başarı sınavı, İngilizce sınavları, konuşma sınavına karşı tutum, yabancı dil olarak İngilizce, Anadolu liseleri.

## ABSTRACT

### **Assessing the Speaking Skill: An Investigation into Achievement Tests of the 9<sup>th</sup> Grade Students in Anatolian High Schools**

Ceyda ÖZDEMİR

Speaking assessment in high schools has recently gathered substantial attention seeing that speaking tests have been compulsory in high schools since 2014 in Turkey (Regulation for Secondary Education Institutions, 2014). This paper offers an investigation of speaking tests in Anatolian high schools. Furthermore, it attempts to describe the whole process of speaking tests in these schools and examine the 9<sup>th</sup> grade students' and their teachers' attitude towards speaking tests. This research was conducted during the spring term of 2016-2017 academic year. Participants were 358 students and 22 English language teachers from six different Anatolian high schools in Denizli. The research combined quantitative and qualitative research methods. Thus, our data were collected through questionnaires and interviews. In the analysis of the data, for questionnaires, Statistical Package for Social Sciences (SPSS) 22.0 and for interviews, content analysis was used. Results revealed that there was no standard process in speaking assessment among schools. Teachers planned speaking test in a detailed way, and then they informed students about testing and grading procedures of speaking tests. In speaking tests, the course teacher assessed each student individually in classrooms where the student's classmates were also. At the end of speaking tests, both teachers and students expressed that they were satisfied with timing. However, teachers pointed out that they gave high marks to show their support and encouragement in scoring, and this condition threatened reliability of the speaking tests. Another finding was that both teachers and students had positive attitudes towards speaking tests. While students were advocating that speaking test was useful to improve their English, teachers believed that they needed it to see the speaking level of students.

**Key words:** Speaking test, achievement test, English language testing, attitude towards speaking test, English as a foreign language, Anatolian high school.

## TABLE OF CONTENTS

YÜKSEK LİSANS TEZ ONAY FORMU.....	iii
ETİK BEYANNAMESİ.....	iv
ACKNOWLEDGEMENTS.....	v
ÖZET.....	vi
ABSTRACT.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xvii
CHAPTER I: INTRODUCTION.....	1
1.1. Background to the Study.....	1
1.2. Statement of the Problem.....	2
1.3. Purpose of the Study.....	3
1.4. Research Questions.....	3
1.5. Significance of the Study.....	4
1.6. Assumptions and Limitations of the Study.....	5
1.6.1. Assumptions of the Study.....	5
1.6.2. Limitations of the Study.....	5
CHAPTER II: LITERATURE REVIEW.....	7
2.1. Theoretical Framework.....	7
2.1.1. Humanism.....	7
2.1.2. Constructivism.....	9
2.1.3. Communicative Language Teaching.....	10
2.1.4. Task-based Language Teaching.....	12
2.1.5. Interlanguage.....	13
2.2. Definitions and Importance of Language Tests.....	13
2.3. History of Testing Speaking as a Second Language.....	15
2.3.1. Historical Background of Teaching English and Testing Speaking Skill in Turkey.....	16
2.4. Positive and Negative Aspects of Speaking Tests.....	19
2.4.1. Listening and Speaking.....	20



2.4.2. Washback.....	20
2.4.3. Reliability.....	21
2.4.3.1. Rater reliability .....	24
2.4.4. Validity .....	24
2.4.4.1. Content validity .....	25
2.4.4.2. Criterion-related validity .....	26
2.4.4.3. Construct validity .....	26
2.4.4.4. Face validity .....	27
2.4.4.5. Response validity .....	27
2.4.5. Practicality .....	27
2.4.6. Objectivity .....	28
2.4.7. Standardization.....	28
2.4.8. Authenticity .....	28
2.4.9. Interactiveness .....	29
2.4.10. Test Anxiety .....	30
2.5. Test Specifications .....	30
2.6. Techniques for Testing Speaking .....	31
2.6.1. Discrete or Integrated Tests? .....	31
2.6.2. Individual or Paired/Group Tests? .....	31
2.6.3. Direct or Indirect Tests? .....	32
2.6.4. Types of Test Tasks for Speaking Assessment .....	33
2.6.4.1. Imitative tasks .....	35
2.6.4.2. Intensive tasks .....	35
2.6.4.3. Responsive tasks .....	36
2.6.4.4. Interactive tasks .....	36
2.6.4.5. Extensive tasks .....	37
2.7. Test Administration .....	37
2.8. Rater and Interlocutor Training .....	38
2.9. The Common European Framework of Reference for Languages (CEFR) .....	39
2.10. Rating Scales .....	41

2.11. Research Studies on Students' and Teachers' Attitudes towards Speaking Tests.....	43
2.12. Conclusion .....	46
CHAPTER III: METHODOLOGY .....	47
3.1. Research Design .....	47
3.2. Setting .....	47
3.3. Participants .....	49
3.3.1. Students .....	49
3.3.2. Teachers .....	49
3.3.3. Participants for the Pilot Study .....	50
3.4. Instruments .....	52
3.5. Data Collection Procedures .....	56
3.6. Data Analysis .....	57
CHAPTER IV: RESULTS .....	59
4.1. The Way How Speaking Skills Are Assessed in Anatolian High Schools .....	59
4.1.1. Planning for the Speaking Test .....	59
4.1.2. Instruction .....	60
4.1.3. Time Allocation .....	62
4.1.4. Test Tasks .....	64
4.1.4.1. Individual or paired/group test tasks? .....	65
4.1.4.2. Discrete or integrated test tasks? .....	67
4.1.5. Materials .....	67
4.1.6. Testing Environment .....	69
4.1.7. Teachers' Roles: Assessor or Interlocutor? .....	71
4.1.8. Teachers' Behaviors .....	73
4.1.9. Objectivity in Assessment .....	74
4.1.10. Rating Scales .....	76

4.1.11. Feedback .....	78
4.2. Students' Perception and Attitudes .....	79
4.2.1. Experience of the Students .....	79
4.2.2. Speaking Tests .....	80
4.2.3. Test Tasks .....	81
4.2.4. Class Activities .....	82
4.2.5. Feelings of the Students .....	84
4.2.6. Feedback .....	85
4.3. Differences between Male and Female Students .....	86
4.4. Differences among Schools in Terms of Students' Placement Scores .....	86
4.5. Teachers' Perception and Attitudes .....	87
4.5.1. Experience of the Teachers .....	87
4.5.2. Speaking Tests .....	88
4.5.3. Test Tasks .....	90
4.5.4. Class Activities .....	90
4.5.5. Feelings of the Teachers .....	91
4.5.6. Rating Scales .....	92
4.5.7. Scoring .....	93
<b>CHAPTER V: CONCLUSION .....</b>	<b>94</b>
5.1. Discussion .....	94
5.1.1. Research Question 1: How are Speaking Skills Assessed in Anatolian High Schools? .....	94
5.1.2. Research Question 2: What are Students' Perception and Attitudes towards the Speaking Test? .....	97

5.1.3. Sub-Question of Research Question 2: Are there any Differences between Male and Female Students? .....	99
5.1.4. Sub-Question of Research Question 2: Are There any Differences among the Schools in Terms of Students' Placement Scores? .....	99
5.1.5. Research Question 3: What are the Teachers' Perception and Attitudes towards the Speaking Test? .....	99
5.2. Pedagogical Implications .....	101
5.3. Suggestions .....	101
REFERENCES .....	103
APPENDICES .....	118
Appendix A: Sample of Holistic Scale .....	119
Appendix B: Sample of Analytic Scale .....	120
Appendix C: Teacher Questionnaire .....	123
Appendix D: Students Questionnaire .....	129
Appendix E: Teacher Interview Questions .....	133
Appendix F: Student Interview Questions .....	134
Appendix G: Teacher Questionnaire (Turkish Version) .....	135
Appendix H: Students Questionnaire (Turkish Version) .....	141
Appendix I: Teacher Interview Questions (Turkish Version) .....	145
Appendix J: Student Interview Questions (Turkish Version) .....	146
Appendix K: Permission from Denizli Directorate of National Education for the Questionnaire .....	147
Appendix L: CV .....	148

## LIST OF TABLES

Table 2.1 The Certificate of Proficiency in English Oral Paper Changes .....	15
Table 2.2 The Descriptors for Overall Spoken Production .....	40
Table 2.3 The Descriptors for Overall Spoken Interaction .....	41
Table 3.1 The Schools Categories in Terms of Students' Placement Scores in Main Study .....	48
Table 3.2 The Schools Categories in Terms of Students' Placement Scores in Pilot Study .....	48
Table 3.3 Age Range of the Students .....	49
Table 3.4 Age Range of the Teachers .....	50
Table 3.5 Experience of the Teachers .....	50
Table 3.6 Age Range of the Students in Pilot Study .....	51
Table 3.7 Age Range of the Teachers in Pilot Study .....	51
Table 3.8 Diversity in the Teachers' Bachelor of Art in Pilot Study .....	51
Table 3.9 Experience of the Teachers in Pilot Study .....	51
Table 3.10 The Items Related to How Speaking skills Are Assessed in Teachers' Questionnaire .....	53
Table 3.11 The Items Related the Teachers' Perception and Attitudes towards the Speaking Test .....	54
Table 3.12 The Items Related to How Speaking Skills Are Assessed in Students' Questionnaire .....	55
Table 3.13 The Items Related the Students' Perception and Attitudes towards the Speaking Test .....	55
Table 3.14 Teacher Interview Questions .....	56

Table 3.15 Student Interview Questions .....	56
Table 4.1 Descriptive Statistics for Planning Process from Teachers' Perspective .....	59
Table 4.2 Descriptive Statistics for Instruction from Teachers' Perspective .....	61
Table 4.3 Descriptive Statistics for Instruction from Students' Perspective .....	61
Table 4.4 Descriptive Statistics for Time Allocation from Teachers' Perspective .....	62
Table 4.5 Test Time from Teachers' Perspective .....	63
Table 4.6 Test Time from Students' Perspective .....	63
Table 4.7 Descriptive Statistics for Time Allocation from Students' Perspective .....	64
Table 4.8 Descriptive Statistics for Test Activities from Students' Perspective .....	64
Table 4.9 Teachers' Preferences for Test Tasks .....	66
Table 4.10 Descriptive Statistics for Test Tasks from Students' Perspective .....	66
Table 4.11 Descriptive Statistics for Test Tasks from Teachers' Perspective .....	67
Table 4.12 Descriptive Statistics for Materials from Teachers' Perspective .....	68
Table 4.13 Descriptive Statistics for Materials from Students' Perspective .....	68
Table 4.14 Descriptive Statistics for Testing Environment from Teachers' Perspective .....	69
Table 4.15 Descriptive Statistics for Testing Environment from Students' Perspective .....	71
Table 4.16 Descriptive Statistics for Teachers' Roles from Teachers' Perspective .....	72
Table 4.17 Descriptive Statistics for Teachers' Roles from Students' Perspective .....	73
Table 4.18 Descriptive Statistics for Teachers' Behavior from Teachers' Perspective..	73
Table 4.19 Descriptive Statistics for Teachers' Behavior from Students' Perspective..	74
Table 4.20 Descriptive Statistics for Objectivity from Teachers' Perspective .....	75

Table 4.21 Descriptive Statistics for Objectivity from Students' Perspective .....	75
Table 4.22 Descriptive Statistics for Rating Scales from Teachers' Perspective .....	77
Table 4.23 Descriptive Statistics for Rating Scales from Students' Perspective .....	78
Table 4.24 Descriptive Statistics for Feedback from Students' Perspective .....	78
Table 4.25 Descriptive Statistics for Students' Experience .....	80
Table 4.26 Descriptive Statistics for Students' Perception and Attitudes Regarding the Speaking Test .....	81
Table 4.27 Descriptive Statistics for Students' Perception and Attitudes Regarding Test Tasks .....	82
Table 4.28 Descriptive Statistics for Students' Perception and Attitudes Regarding Class Activities .....	83
Table 4.29 Descriptive Statistics for Students' Perception and Attitudes Regarding Students' Feelings .....	84
Table 4.30 Descriptive Statistics for Students' Perception and Attitudes Regarding Feedback.....	85
Table 4.31 Attitude of Male and Female Students towards Speaking Tests .....	86
Table 4.32 Descriptive Results of One-Way ANOVA in Terms of School Categories.....	87
Table 4.33 Attitude of the Schools towards Speaking Tests .....	87
Table 4.34 Descriptive Statistics for Teachers' Experience .....	88
Table 4.35 Descriptive Statistics for Teachers' Perception and Attitudes Regarding the Speaking Test .....	89

Table 4.36 Descriptive Statistics for Teachers' Perception and Attitudes Regarding Test Tasks .....	90
Table 4.37 Descriptive Statistics for Teachers' Perception and Attitudes Regarding Class Activities .....	90
Table 4.38 Descriptive Statistics for Teacher Preferences for the Frequency of Speaking Tests .....	91
Table 4.39 Descriptive Statistics for Teachers' Feelings .....	92
Table 4.40 Descriptive Statistics for Teachers' Perception and Attitudes Regarding Rating Scales .....	92
Table 4.41 Descriptive Statistics for Teachers' Perception and Attitudes Regarding Scoring .....	93



## LIST OF FIGURES

Figure 2.1 Potential factors affecting testing process .....	22
Figure 2.2 Test taker characteristics that affect students' performance .....	23
Figure 2.3 Informational tasks .....	33
Figure 2.4 Interactional tasks .....	33
Figure 2.5 Types of tasks in speaking assessment .....	34
Figure 2.6 Types of assessment tasks in speaking skill.....	35
Figure 2.7 Activities of spoken production and interaction .....	40
Figure 3.1 The demonstration of gender of the students .....	49
Figure 3.2 The demonstration of gender of the teachers .....	50
Figure 4.1 The percentage of the teachers who plan timing before a speaking test.....	59
Figure 4.2 The demonstration of the speaking tests' frequency in a term .....	60
Figure 4.3 The demonstration of the materials usage during the speaking test .....	68
Figure 4.4 The demonstration of the teachers' preferences for the rating scales .....	77
Figure 4.5 The percentages of the students who experience speaking tests .....	80
Figure 4.6 The percentage of the teachers who do interlocutor training .....	88

## CHAPTER I: INTRODUCTION

This chapter contains six sections: background to the study, statement of the problem, purpose of the study, research questions, significance of the study, assumptions and limitations of the study. The first is background to the study. In this section, theoretical framework of the testing speaking will be presented in general. The second is statement of the problem. In this section, problems of the speaking tests will be highlighted. Next, purpose of the study will be pleaded; research questions will be notified afterwards. Then, significance of the study will be specified. Finally, assumptions and limitations of the study will be clarified.

### 1.1. Background to the Study

Constructivism is a powerful learning theory. It is against traditional approach in which teachers pass on knowledge solely and students just accept the knowledge in a passive way. In contrast, in constructivism, students construct their own knowledge in the learning process. Constructivism supports student-centered learning. As such, students have a part in learning and teaching process actively. In the case of foreign language learning, students have to speak English in order to develop their speaking skill.

Speaking skill has not been taken to forefront in the major dominant approaches like grammar-translation method and audio-lingual method. Grammar-translation method ignores speaking skill, and audio-lingual method also lacks fluent, spontaneous, native-like speech although it is proficient in supplying accuracy (Hall, 2011). Dissatisfaction for these methods pushes into a new approach: communicative language teaching (CLT), and speaking ability rises to prominence.

CLT is seen as “the most influential approach in the history of second/foreign language instruction” (Spada, 2007, p. 283). Inasmuch as, according to Richards and Rodgers (2014), CLT “aimed to (a) make communicative competence the goal of language teaching and (b) develop procedures for the teaching of the four language skills that acknowledge the interdependence of language and communication” (p. 85). However, teaching English has been mostly about grammar, reading and vocabulary, which are often based on accuracy so far. With the advent of CLT, language is seen as a tool for communication.

Task-based language teaching (TBLT) is another approach which clears the way for speaking skill. This method gives priority to students’ performance of tasks. Students need to complete tasks by communicating actively. Moreover, learners have already needed to

speak English for many purposes such as their education, professions, travel and interests. Eventually, speaking skill cannot be ignored anymore. Teachers need to create opportunities and urge their students to use English as much as possible.

With the emergence of these notions, assessing the speaking ability has become one of the important issues in language testing (Sak, 2008; Morrow, 2012). Speaking tests provide information about learners' progress and their needs. Çaykan (2001) points out that "the students tend to attach more importance to developing oral skills when they expect an oral test than when they do not" (pp. 14-15). Paker (2013) also reiterates this idea and states that students cannot ignore any skill as long as it is assessed. In addition, Ur (1996) reports that the introduction of speaking tests in Israel has drawn more attention towards speaking skill in schools.

In Turkey, speaking tests have been compulsory in high schools since 2014 (Regulation for Secondary Education Institutions, 2014). In essence, speaking tests in high schools are the outcome of innovations and reforms in English language curriculum. The updated curriculum, which is based on communicative competence, upholds none of skills can be passed over in the process of language learning (Ministry of National Education, 2013; Ministry of National Education, 2014).

## **1.2. Statement of the Problem**

Assessing speaking skill is difficult in many ways. As such, trying to take the most appropriate, valid and reliable decisions about the construction, administration, scoring and interpretation of a speaking test makes it clearly challenging for test designers. It is also challenging for test takers "because speaking is done in real-time, learners' abilities to plan, process and produce the foreign language are taxed greatly" (Luoma, 2004, p. ix). Furthermore, "the nature of the interaction, the sorts of tasks presented to the candidate, the questions asked..., and the opportunities provided to show his or her ability to speak in a foreign language" all influence the performance of the test taker (Luoma, 2004, p. x).

Sak (2008) also exemplifies some problems of speaking tests: administrative costs, difficulties of testing a large number of students individually or in small groups, training teachers and the total amount of time and the number of teachers needed for administering the tests. In addition to all these, J.D. Brown (1996) deals with scoring problem as follows: "the subjective nature of the scoring procedures can lead to evaluator inconsistencies or shifts having an effect on students' scores and affect the scorer reliability adversely" (p. 191). The nature of the interaction, the test methods, the topics, the interlocutor effect and

test taker characteristics give reason for possible variability in speaking test scores (A. Brown, 2003; A. Brown, 2005; Berry, 2007; Brindley, 1991; O’Sullivan, 2006; Shohamy, 1988; Shohamy, 1994). Despite the problems, it is possible to minimize or overcome them by developing testing procedures attentively, constructing tasks and rubrics carefully and training raters continuously.

World-renowned educational organization Education First English Proficiency Index (2017) has ranked countries by their English skills since 2011. According to its 2017 announcement, Turkey has very low English proficiency. Furthermore, Turkey is ranked 62<sup>nd</sup> among 80 countries in the world and second to last among the 27 European countries included in the survey. In 2012, Paker conducted a research about why we cannot teach foreign language and why our students cannot learn English communicatively. According to this research, 95% of the English language teachers tries to teach grammar as a priority in Turkey, and they consider that English is for teaching, not for communication. This attitude is also reflected in the tests in which there are no listening, writing and speaking sections. Although some tests include only reading parts, most sections are usually based on grammar and vocabulary. Accordingly, students try to learn just grammar and vocabulary in order to pass the tests but ignore four language skills as negative washback. Initially, students are generally eager to learn English when they start to learn it; however, they lose their motivation in progress of time, and even they may hate English because they think that they can never learn it due to grammar focused teaching and testing activities. Policy makers have tried to change this sense and reformed English language curriculum and regulation.

### **1.3. Purpose of the Study**

The aim of this study is to identify the whole process of speaking tests with regards to planning phase of the test, instruction, time allocation, test tasks, materials, testing environment, teachers’ roles and behavior, objectivity, rating scales and feedback of English language teachers in these schools. The study is also to reveal perception and attitudes of the English language teachers and the 9<sup>th</sup> grade students towards speaking tests in Anatolian high schools. Moreover, the study intends to research whether students’ attitudes differ by their gender and schools.

### **1.4. Research Questions**

This study addressed the following research questions:

- 1) How are speaking skills assessed in Anatolian high schools?
- 2) What are students’ perception and attitudes towards the speaking test?

- a) Are there any differences between male and female students?
  - b) Are there any differences among the schools in terms of students' placement scores?
- 3) What are the teachers' perception and attitudes towards the speaking test?

### **1.5. Significance of the Study**

Majidifard, Shomoossi & Ghourchaei (2014) describe the critical role of speaking a foreign language and indicate that speaking is the prominent skill for language learners due to globalization and the widespread use of English in the world today. Accordingly, research on speaking has been on the rise in the EFL context. Learning a foreign language is composed of many skills such as reading, writing, speaking and listening, grammar, vocabulary. However, as Lazaraton (2001) points out, people consider that knowing a language is to be able to speak it. If they cannot speak the language, why do they learn it? Why do they spend their time, effort or money for it? All in all, “of all the four skills, speaking seems intuitively the most important: people who know a language are referred to as speakers of that language” (Ur, 1996, p. 120).

Speaking skills are an indispensable part of the curriculum in language teaching, and this makes them an essential object of assessment as well (Luoma, 2004). Therefore, policy makers amended regulation in high schools. Since 2014, speaking tests have been compulsory in high schools (Regulation for Secondary Education Institutions, 2014). The regulation regarding the assessment of English clarifies that “the examinations of language courses are conducted as pencil-and-paper tests and performance tests to measure listening, speaking, reading and writing skills” (Regulation for Secondary Education Institutions, 2014, Article 45/1-h). Previously, the four English language skills were not assessed equally in schools. Listening, speaking and writing skills were ignored in testing while grammar, vocabulary and reading skills were being emphasized.

Speaking skill has been clearly assessed in many institutions and universities as a choice of teachers or institutions. With the new regulation, all high schools have to assess their students' speaking skill (Regulation for Secondary Education Institutions, 2014). Even so, there is no clear format for speaking tests developed and imposed by Ministry of National Education. There are just some suggestions like discussion time activities or video blogs (V-logs) for speaking tests in the 9th-12th grades English curriculum (Ministry of National Education, 2014). In the regulation, “...the type, the number and timing of tests are determined by teachers, they are carried out depending on the schoolmaster's approval”

(Regulation for Secondary Education Institutions, 2014, Article 45/1-f). Consequently, each high school chooses its own way.

In my study, I attempted to find out how teachers assessed speaking skill in high schools. The findings of the study may be a trigger for other teachers who deal with similar problems concerning speaking test. They can compare their way of testing with the findings of this study. In this regard, either they may keep going on what they do or they may make some changes in their testing. In this study, I also wanted to investigate perception and attitudes of teachers and students toward speaking assessment in high schools because they have had an impact upon teaching/learning process, and what is more, the ultimate success or failure of speaking assessment.

In my opinion, this study will contribute to the EFL field in Turkey and will be beneficial for the ones who want to study in this field. There seems no prominent research reflecting what perception and attitudes of teachers and students or how speaking tests are applied in high schools although there have been few studies like Güllüoğlu (2004), Höl, (2010), Duran (2011) on perception and attitudes of teachers and students in universities.

## **1.6. Assumptions and Limitations of the Study**

### **1.6.1. Assumptions of the Study**

This study assumes that:

- The students and the teachers are sincere as responding to the questionnaires and interview questions.
- Anatolian high schools in this study arrange a speaking test each semester.
- Teachers use some rubrics for the speaking test.
- Students in the 9<sup>th</sup> grade have similar language level.
- The teachers and the students may be inexperienced about testing speaking, and they may have difficulties during the test because there is no obligation for speaking tests before high school.

### **1.6.2. Limitations of the Study**

This study is carried out in spring semester of 2016-2017 academic year. It is only limited to six Anatolian high schools in Denizli. Thereby, it is a local study and is not generalized to other high schools such as science high school, social sciences high school, and vocational and technical Anatolian high school around the country. Obviously, there is a need for research on other types of high schools.

358 students and 22 teachers participated in the study. More participants could be in this study. Furthermore, all students were in the 9<sup>th</sup> grade in Anatolian high schools.

Therefore, they generally had similar level (the CEFR A1 & A2). If the participants were different levels, this study would be more valid and reliable. Eventually, there is a need for research on other graders like the 10<sup>th</sup>, 11<sup>th</sup> and 12<sup>th</sup>.

## **CHAPTER II: LITERATURE REVIEW**

In this chapter, theoretical framework, definitions and importance of language tests, history of testing speaking as a second language, positive and negative aspects of speaking tests, test specifications, techniques for testing speaking, test administration, rater and interlocutor training, The Common European Framework of Reference for Languages (CEFR), rating scales, and research studies on students' and teachers' attitudes towards speaking tests will be reviewed.

First of all, humanism, constructivism, communicative language teaching, task-based language teaching and interlanguage will be explained as theoretical framework. Secondly, definitions and importance of language tests will be underlined. History of testing speaking as a second language will be revealed subsequently. Later, positive and negative aspects of speaking tests will be discussed. Next, test specifications will be depicted. After that, techniques for testing speaking will be identified. Besides, procedures for test administration will be characterized. Afterwards, rater and interlocutor training will be represented. The Common European Framework of Reference for Languages (CEFR) will be analyzed. Then, rating scales will be interpreted. Lastly, research studies on students' and teachers' attitudes toward speaking tests will be outlined.

### **2.1. Theoretical Framework**

Speaking tests shine out with respect to some notions and approaches like humanism, constructivism, Communicative Language Teaching (CLT), Task-based language teaching (TBLT) and interlanguage. All these notions support student-centered approach. As such, students participate in learning process actively, and individual needs of the students are put at forefront. In evaluation process, tests are tools showing each student's progress. In this sense, tests help students to gain awareness relevant to their strengths and weaknesses. Pencil-and-paper tests are not effectual to assess students' speaking skills. In speaking tests, students need to perform their abilities in authentic and communicative situations for reliability and validity of the tests.

#### **2.1.1. Humanism**

Humanistic approach emerged in the mid-20th century (Vasuhi, 2011). It was founded on works of Abraham Maslow (1962) and Carl Rogers (1965). Humanism gives priority to the importance of the individual and specific human needs. The major assumptions of humanism are (a) human nature is inherently good; (b) individuals are free and autonomous, hence they can make major personal choices; (c) human beings possess unlimited potential for growth and development; (d) self-concept has a critical role in



growth and development; (e) individuals have an urge toward self-actualization; (f) reality is defined by individually; and (g) individuals have responsibility to both themselves and to others (Elias and Merriam, 1980).

Humanistic education focus on human well-being, including the primacy of human values, ideas, opinions, feelings, interests, goals and experiences; the realization and development of human potential; and the recognition of human dignity over any other economic, religious, nationalistic or ideological set of values (Aloni, 1997; Gadamer, 1975; Sharp, 2012). The primary goal of humanistic education is to facilitate development of self-actualizing persons (McKenna, 1995; Patterson, 1973). Self-actualizing is the process of realizing and expressing one's own capabilities and creativity (Khatib, Sarem & Hamidi, 2013). According to G. Wang (2005), if a person cannot satisfy his/her basic needs physically and psychologically, s/he might fail to concentrate on his/her language learning heartily. So “foreign language teachers must contribute to the self-actualizing process” (Medgyes, 1986, p. 109).

Humanistic education is a student-centered approach in which students engage in learning process actively. Teachers and students designate learning preferences and evaluate learning processes co-operatively. Moskowitz (1978) maintains that each student is unique, and each one has a different personality, feeling and interest. According to Mishra (2000), humanistic teachers trust and see students as valuable individuals. The responsibility of teachers is to identify students' needs and serve as a facilitator, not knowledge transmitter. As A. Underhill (1989) points out, “the facilitator has a lot to do with setting the mood or atmosphere which supports self-directed learning” (p. 256). In the created atmosphere, the students are supposed to use the target language in meaningful contexts for real functions such as expressing their own thoughts and feelings. Interaction increases students' motivation and sociability. After all, facilitation occurs only through interpersonal relationship that encourages friendship and cooperation, then accelerates language learning.

Humanism stresses the inner world of humans and views the thoughts and feelings of individuals as the foreground of other human achievements (G. Wang, 2005). According to Moskowitz (1978), “humanistic education takes into consideration that learning is affected by how students feel about themselves” (p. 12). In other words, the more students think or feel positively, the more they are facilitated to learn. Thus, the development of a positive self-concept leads to students' achievement (Khatib, Sarem & Hamidi, 2013). Only when students' self-esteems are raised, do they take responsibility for their own learning as autonomous students (Sharp, 2012) because it is a motivating factor to increase

students' self-esteem in learning (Khatib et al., 2013). Humanistic assessment must highlight the students' strengths, needs and expectations to motivate them (De Matos, 2005). If assessment is applied in a negative way, it will be demotivating for the students. Teachers should take knowledge and feelings into consideration as evaluating students because intellect and emotion are interlinked.

### **2.1.2. Constructivism**

Constructivism roots in the cognitive theories of Piaget (1955) and in the socio-cultural theory of Vygotsky (1978). According to constructivists, reality is constructed personally, namely, personal experiences determine reality (Cooper, 1993). While Piaget (1955) asserts that reality is constructed individually, Vygotsky (1978) upholds that reality is constructed socially. However, they do not deny absolute realities; just combine them with personal or social experiences because realities proceed from interpretations of the experiences.

The main contribution of constructivism is the student-centered approach. In constructivism, the teachers' role is not knowledge transmitter anymore. Students also do not receive knowledge by memorizing or repeating, instead they construct their own knowledge by thinking, understanding, applying and analyzing as a part of active learning (Marlowe & Page, 2005). Each student has their own unique background knowledge and experience so they learn individually. Teachers encourage students to ask critical questions, share their experiences, and exchange knowledge interactively (J.G. Brooks & M.G. Brooks, 1993). Hence, "students and teachers play a role in facilitating and generating knowledge" (Aljohani, 2017, p. 106).

Teachers in constructivism (a) embed learning in complex, realistic and relevant learning environments; (b) provide for social negotiation and shared responsibility as a part of learning; (c) support multiple perspectives and use multiple representations of content; (d) nurture self-awareness and an understanding that knowledge is constructed; and (e) encourage ownership in learning (Driscoll, 2000).

Constructivists advocate the autonomy of students (Aljohani, 2017). Students need to become aware of their learning style and take responsibility of their own learning strategies and techniques applied during lesson. At this point, teachers as facilitators assist students to understand their own cognitive processes and form their own learning awareness. All in all, students are "active, self-regulating, reflective" in the process of learning (Seels, 1989, p. 14).

Wilson (1996) describes a constructivist learning environment as “a place where learners may work together and support each other as they use a variety of tools and information resources in their guided pursuit of learning goals and problem-solving activities” (p. 5). The environment should be supportive, nonthreatening for cooperative learning. Students are motivated intrinsically through authentic tasks and materials. Thus, they participate in critical thinking, problem solving and authentic activities which they use target language creatively to accomplish by using of relevant abilities (O’Malley & Valdez-Pierce, 1996).

According to constructivists, assessments should have more of a 'real-life' application (Herman, Aschbacher and Winters, 1992). In addition, they prefer more interactive and experience based assessments such as authentic, performance, or portfolio assessment (Rami, Lorenzi & Lalor, 2009). These kinds of assessments are stimulating form of evaluation. Constructivists do not view assessment as an outcome, but rather an ongoing process that helps the students continue to learn (Holt & Willard-Holt, 2000) by judging their own progress or each other’s progress (J.G. Brooks and M.G. Brooks, 1993).

### **2.1.3. Communicative Language Teaching**

Communicative language teaching (CLT) was proposed in the early 1970s in the UK and at the start of the 1980s in the US (East, 2016). A huge desire for language learning came into existence on account of the economic development and widespread migration (Xiaotong, 2014). Ultimately, the former methods could not cover demands. The introduction of CLT induced a significant shift in pedagogy away from a linguistic emphasis to communicative emphasis (East, 2016). In CLT, “what it means to know a language and to be able to put that knowledge to use in communicating with people in a variety of settings and situations” (Hedge, 2000, p. 45).

Nunan (1991) explains five features of CLT:

- an emphasis on learning to communicate through interaction in the target language,
- the introduction of authentic texts into the learning situation,
- the provision of opportunities for learners to focus, not only on language but also on the learning management process,
- an enhancement of the learner's own personal experiences as important contributing elements to classroom learning
- an attempt to link classroom language learning with language activities outside the classroom (p. 78).

As an inference from the features, CLT is a learner-centered approach which capitalizes on the interests and needs of the learners. Moreover, CLT is based on humanistic approach so learners are responsible for their own learning. Learners engage meaningful and authentic language use as negotiator and interactor in classroom activities. Teachers are active facilitators of the communication process, and they talk less and listen

more. Both teachers' and learners' motivation and positive attitude are crucial for effective teaching and learning.

Savignon (1972) claims that the primary goal of CLT is to develop learners' communicative competence in interactive situations of real life. Communicative competence emphasizes the use of target language for communication. First, Hymes (1972) introduced this term to linguistic discourse both as an inherent grammatical competence and the ability to use it in a variety of communicative situations. Then, many linguists such as Savignon (1972), Canale and Swain (1980), Canale (1983), Widdowson (1983), Bachman and Palmer (1996), The Common European Framework References (CEFR) (Council of Europe, 2001), and H.D. Brown (1994) give their valuable contribution to the further development of the concept of communicative competence (Bagarić & Djigunović, 2007). Nevertheless, Canale and Swain's model is dominant by reason of its easiness (Bagarić & Djigunović, 2007).

Canale and Swain (1980) want to "establish a clear statement of the content and boundaries of communicative competence... that will lead to more useful and effective second language teaching, and allow more valid and reliable measurement of second language communication skills" (p. 1). According to the model, there are four genres of competences: grammatical, sociolinguistics, strategic and discourse competences (Canale, 1983).

- ❖ Grammatical competence encompasses phonology, morphology, semantics, lexis, grammar and orthographic rules (Council of Europe, 2001). Students use forms and rules of language for communicative purposes and express themselves accurately.
- ❖ Sociolinguistic competence is concerned with possession of knowledge and skills to use appropriate language in various socio-cultural contexts. For example, language elements marking social relationships, rules of appropriate behavior, and expressions of people's wisdom, register differences, style, dialects, accent, stress, and so on.
- ❖ Strategic competence deals with the use of verbal and non-verbal communication strategies to compensate for breakdowns in communication owing to insufficient competence or performance limitation and to enhance the efficiency of communication (Canale, 1983). These strategies include paraphrase, circumlocution, repetition, reluctance, avoidance of words, structures or themes, guessing, changes of register and style, modifications of messages, etc.

- ❖ Discourse competence refers to the ability to combine language structures into different types of cohesive (meaning) and coherent (form) texts (e.g. letter, political speech, poetry, academic essay, cooking recipe).

All these competences turn out and form communicative competence. It is impossible to break them apart, so all competences are equally worthy. Savignon (1972) claims that competence can be observed, developed, maintained and evaluated only through performance. As such, CLT has had significant implications for assessment. Pencil-and-paper tests are no longer sufficient for communicative tests (Clark, 1972; J.B. Carroll, 1961; Jones, 1977; Morrow, 1977; Oller, 1976). In communicative testing, students demonstrate their knowledge in a meaningful and authentic communicative situation (Fulcher & Davidson, 2007). Bachman (2000) suggests that teachers should “take into consideration the discursal and sociolinguistic aspects of language use, as well as the context in which it takes place” (p. 3) to assess language proficiency.

#### **2.1.4. Task-Based Language Teaching**

Task-based language teaching (TBLT) was first developed by Prabhu (1987) in India. He believes that students can learn more effectively when their minds are focused on the task, rather than on the language they are using. A task means “an activity which requires learners to use language, with emphasis on meaning, to attain an objective” (Bygate, Skehan and Swain, 2001, p. 11). Tasks will foster effective language acquisition since students participate in authentic and meaningful communicative situations (D. Willis & J. Willis, 2007; East, 2015; Nunan, 2004). To sum up, TBLT is an approach that offers students plentiful and varied opportunities to actively engage in real communication so as to achieve a goal or complete a task (D. Willis & J. Willis, 2007).

TBLT has its origins in communicative language teaching, yet it constitutes a strong version of CLT (Skehan, 2003). On the grounds that, TBLT aims primarily fluency even though it does not neglect accuracy (East, 2012; H.D. Brown, 1994). Like CLT, TBLT is a student-centered approach and respects students’ interests (Breen, 1987; Long, 1985; Prabhu, 1987). Teachers monitor task process, and they select varied tasks that respond to potentially diverse learner types and need to motivate students for a meaningful communication (Norris, 2009).

Tasks present learners with information gaps to be overcome, problems to be solved, decisions to be made, or otherwise meaningful reasons to interact with each other, that negotiation of meaning often leading to communication breakdown and the opportunity for self-, other-, or teacher-initiated feedback (Norris, 2009, pp. 583-584).

Students may analyze what they do, what works, and what does not by the virtue of feedbacks, thereby, they construct their own explanations such as a close understanding of

their development and a clear target for learning (Dewey, 1938). In this sense, tasks are motivational for students, and they raise their awareness of learning process. Task-based assessment also emphasizes the performance of target tasks, rather than the demonstration of the language knowledge (Norris, 2009; Wiggins, 1998).

### **2.1.5. Interlanguage**

Interlanguage is a term introduced by Larry Selinker in 1972. It is a language of the learners in the process of learning. This language is between target language and native language, but it is like neither native language nor target language. According to Selinker (1972), interlanguage refers to an entirely new language system that learners construct unique to themselves at any stage of learning. They also revise this system in orderly and predictable ways over time (Ellis, 2008). *Processability theory* attempts to understand how learners' restructure their interlanguage systems (Pienemann, 1998). Clahsen (1984) sets forth that certain processing principles determine this order of restructuring. Specifically, he points out that learners first, maintain declarative word order while changing other aspects of the utterances, second, move words to the beginning and end of sentences, and third, move elements within main clauses before subordinate clauses. A. Hughes (1983) suggests several conditions under which learners edit their interlanguage:

- the learners continue to have unsatisfied communicative needs,
- the learners continue to communicate,
- the situations in which they attempt to communicate are sufficiently frequent,
- the learners understand at least some of the language when they communicate.

It is assumed that the learners' native language affects their learning a target language positively or negatively. Inasmuch as learners use their knowledge of native language to understand or produce meaning in target language. It is called as *language transfer*. To illustrate, Tarone (1982) assert that interlanguage speakers have a vernacular style no matter how advanced speakers. In the process of learning target language, learners make numerous errors due to native language interference. But according to Harmer (2003), errors are a part of the learners' interlanguage. In other words, errors are considered as a reflection of the learners' temporary language system. Therefore, they are not signs of failure but a natural part of the learning process (Corder, 1981). Teachers should be conscious of this process while teaching and testing English.

## **2.2. Definitions and Importance of Language Tests**

Language tests are instruments that “provide an accurate measure of the test-taker's ability within a particular domain” (H.D. Brown, 2004, p. 4). This definition sounds fairly simple. On the other hand, the following definitions dwell on learning and teaching process

of English because language tests have a crucial role in each stage of learning and teaching (Douglas, 2014; Hosseini & Azarnoosh, 2014). For example, Carey (1988) reiterates this interpretation with this statement: “language testing is an integral part of the teaching and learning process, and it provides teachers with vital information” (p. xv). Bachman and Palmer (1996) also interpret language tests as “a valuable tool for providing information that is relevant to several concerns in language teaching” (p. 8). Sheng-ping & Chong-ning (2004) define tests as “the main sources for both teachers and students to get feedback, which enables them to reflect on their teaching and learning activities and thus to help improve the jobs of both sides” (p. 1).

Madsen (1983) asserts that “language tests can benefit students, teachers and even administrators by confirming progress” (p. 5), and he (1983) clarifies that language tests give students a sense of accomplishment and a feeling that the teacher’s evaluation of them matches what he has taught them. As a consequence, this creates positive attitudes towards instruction. However, students can experience a sense of failure, and language tests can present a challenge for some students. By all means, it can be discussed whether this challenge is motivating or not. But Madsen (1983) sets forth that language tests show students what they need to improve in this circumstance, thus students can study hard according to their needs.

Furthermore, language tests help teachers to

make decisions about individuals, such as determining what specific kinds of learning materials and activities with regard to instructional objectives should be provided to students, based on diagnosis of their strengths and weaknesses, deciding whether individual students or an entire class are ready to move on to another unit of instruction, and assigning grades on the basis of students’ achievement (Bachman and Palmer, 1996, p. 8).

To sum up, teachers need language tests as a tool for providing a feedback on the effectiveness of the teaching program. Hence, they can organize their future efforts and testing plans in connection with teaching and learning process under the guidance of feedback. In other words, a test can improve evaluation process itself such as preparing, administering, timing, scoring, validity and reliability of tests because a test is upgraded as long as it is applied.

Douglas (2014) yields that “tests also allow other stakeholders, including programme administrators, parents, admissions officers and prospective employers, to be assured that learners are progressing according to some generally accepted standard or have achieved a requisite level of competence in their second language” (p. 1). When all these benefits are considered, they raise the importance of language tests.

### 2.3. History of Testing Speaking as a Second Language

“The theory and practice of testing second language speaking is the youngest subfield of language testing” (Fulcher, 2014, p. 1). However, the assessment of speaking skill has evolved dramatically over the last several decades from test of oral grammar and pronunciation to interviews, and more recently, to multiple tasks, often collected over time (Cohen, 1994). Nowadays, it is so popular to assess speaking skill in communication-oriented tasks in which language learners “structure information effectively and communicate smoothly in a socially acceptable manner” (Luoma, 2004, p. 187).

Speaking tests in the United States actually got the center of attention during the Second World War (Fulcher, 1997). Beforehand, “testing speaking was frequently seen as desirable but not feasible” because of problems with reliability and practicality (Fulcher, 2014, p. 5). Fulcher (2014) deduces that “political and military events have had a deep impact upon the form and scoring of many modern speaking tests” (p. 1) because soldiers needed second language to carry out their duty effectively, hence they would not have communication problems. Testing system for speaking skill in military was updated and exported to schools in the later years.

Unlike in the United States, the primary purpose of speaking tests in the United Kingdom was to support the syllabus and encourage good teaching and learning (Brereton, 1944). Roach (1945) revealed that modern speaking tests started with the first general proficiency examination, published by the University of Cambridge Local Examination Syndicate in 1913. Weir (2003) tracks the changes to the Certificate of Proficiency in English (CPE) speaking since its introduction, and they are summarized in Table 2.1.

Table 2.1

#### *The Certificate of Proficiency in English Oral Paper Changes*

CPE	Task	Task	Task	Task	Assessment
1913	Dictation	Reading Aloud	Conversation		Unclear
1934	Same	Same	Same		Unclear
1938	Same	Same	Same	Discretionary task (story dictation)	Unclear
1945	Same	Same	Same		Unclear
1953	Same	Same	Same		Unclear
1966	Same	Same	Same		Unclear
1975	Interview Questions based on photograph	Long turn (topic given 15 minutes before test – notes allowed)	Reading aloud – extract from play given 10 minutes before test. Jointly read by candidate and	Listen to passage read by examiner and give appropriate response	Rating Scale: Vocabulary, Grammar & Structure, Intonation, Rhythm,



			examiner		Stress & Pronunciation, Overall communication
<b>CPE</b>	<b>Task</b>	<b>Task</b>	<b>Task</b>	<b>Task</b>	<b>Assessment</b>
1984 *	Same	Similar but this task combined with old task 4	Same but candidate only reads short passage	Roleplay	Fluency, Grammatical Accuracy, Pronunciation & Stress, Communicative Ability, Vocabulary
2003 †	1-to-1 interview	Paired/small group discussion based on photographs	Individual long turn	Group Discussion	Grammatical & Lexical Resource, Discourse Management, Pronunciation

\* Provision for groups and individual tasks marks using selected criteria

† Paired format (groups of three possible)

As can be seen in Table 2.1, there was no clear assessment form before 1975, and the tasks had in earnest limitations. For example, the dictation task is “a combination of listening and writing skills rather than speaking, while the reading aloud task engages low level reading skills together with pronunciation and intonation control” (O’Sullivan, 2013, p. 266). The modernization of the CPE oral paper really began with the 1975 version. The time coincides with communicative language teaching (CLT). Parallel to the development of CLT, communicative language testing has gained popularity as an assessment tool for oral proficiency in 1970s and 1980s. Techniques such as pair and group work, task-based learning, the concept of language awareness all appeared after this method was introduced (R. Hughes, 2002). The transition from interview to pair or group testing was claimed to have positive washback on the classroom in terms of time saving. (Egyud and Glover, 2001; Ducasse and Brown, 2009).

### 2.3.1. Historical Background of Teaching English and Testing Speaking Skill in Turkey

The status of English as an international language has been acknowledged for several decades (Crystal, 1997; Kachru, 1992; K. Brown, 2002; Widdowson, 1994). What is more, Seidlhofer (2001) claims that English should be considered as a lingua franca (ELF) when it is used for international communication. In Turkey, English is also seen as a lingua franca, an international language and the language of science and technology (Ministry of National Education, 2014). English has not got any official status but

enormous prestige in Turkey. Turkish people believe that they will gain access to better education and a more prestigious job with good benefits and prospects for promotion (Kırkgöz, 2005b; Kızıltepe, 2000), and they can follow technological and scientific improvements easier by the virtue of English knowledge (König, 1990).

English was introduced to Turkish education system in Tanzimat Period during the second half of the eighteenth century when the movements for westernization started (Kırkgöz, 2007). English was needed in an effort to sustain communication with the outside world for economic, social, and business relations to accelerate Turkey's modernization and westernization process (Demirel, 1990). Grabe (1988) also reiterates this view with this statement "any country wishing to modernize, industrialize, or in some way become technologically competitive, must develop the capacity to access and use information written in English" (p. 65). Accordingly, English language gained admission, at first, in the Ottoman military schools, then, in the public schools (Boyacıoğlu, 2015).

In 1776, Mühendishane-i Bahrî-i Hümâyûn [The Imperial Maritime Engineering Schools] was the first military educational school, using English as a medium of instruction because teachers were English in this school (Ergün, 2000). After that, Robert College, founded as the first private English-medium secondary school by an American missionary in 1863, and Galatasaray Sultanisi, founded as the first State school to teach in foreign language by Ottoman Empire in 1868, offered English courses (Kırkgöz, 2007; Özbay, n.d.).

With the foundation of Turkish Republic in 1923 after the decline of Ottoman Empire, a closer connection with Europe and the USA accelerated the spread of teaching English (Kırkgöz, 2007). Maarif colleges were founded with the purpose of teaching English in Turkey (Çetintaş & Genç, 2001). The first one was Yenişehir Lisesi, founded in 1932 (Demircan, 1988). Maarif colleges have been known as Anatolian high schools since 1975 (Encyclopedia Britannica, 1993).

Anatolian high schools were seven-year state schools that taught some of the courses like Mathematics and Science in English, German or French, assigned the students with a central examination system, applied Middle school (Basic Education) and high school curriculum with one academic year preparatory classes (Special Regulation for Anatolian High Schools Which Teach Some Lessons through a Foreign Language, Galatasaray High School - İstanbul Erkek High School, 1976). The aim of the foreign language teaching in these schools was identified as "to enable the students to speak and understand the foreign language they study, to translate their texts into Turkish and to

express it sufficiently in writing" (Regulation for Higher Education Institutions Which Teach a Foreign Language and Teach through a Foreign Language, 1984, p. 4). Anatolian high schools had a distinct status among the other public schools, and they were seen as the golden key to access prestigious universities and thus a prosperous future (G.Sarıçoban & A.Sarıçoban, 2012). However, the interest in these schools is diminishing as the number of these schools is increasing. They have lost their success pin nowadays.

Globalization has also brought about an unprecedented spread of English in 1980s (Friedman, 1994; Kırkgöz, 2009; Robins, 1996). Ultimately, Foreign Language Teaching and Learning Act issued in 1983 (Foreign Language Teaching and Learning Act, 1983). Thus, foreign language teaching, English language curriculum and syllabi to be followed in schools were standardized by Turkish Ministry of National Education.

Up to now, Turkish government has frequently changed English teaching policy. According to needs of teachers and students, many reforms and innovations have taken place in Turkish education system in terms of teaching, testing, curriculum and syllabi. Since 1997 reform, English language curriculum has been consonant with the language teaching standards of the European Union (Kırkgöz, 2009). Accordingly, the reforms are based on the principles and descriptors of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR).

The CEFR particularly stresses the need for students to put their learning into real-life practice so as to support fluency, proficiency and language retention (Council of Europe, 2001). The updated curriculum was designed to take all aspects of communicative competence into consideration in an authentic communicative environment instead of focusing on the language as a topic of study by addressing functions and four skills of language in an integrated way (Ministry of National Education, 2013; Ministry of National Education, 2014). The CEFR promotes student-centered learning in order to replace the traditional teacher-centered view, and it supports learners in becoming language users rather than students of the language (Council of Europe, 2001). The main goal is to make learners effective, fluent, and accurate communicators in English (Ministry of National Education, 2013).

After all, the CEFR has not been prescriptive in the implementation of these innovations (G.Sarıçoban & A.Sarıçoban, 2012). It is still criticized that the English classes lack effective communicative competence with too much focus on teaching and assessing grammatical structures (Ministry of National Education, 2014). Bamgbose (2003) points out "no matter how desirable language policies may be, unless they are backed by the will to implement them, they cannot be of any effect" (p. 428). In light of this statement,

teachers as policy makers in practice have crucial responsibilities for implying the policy issues into practice (G.Sarıçoban & A.Sarıçoban, 2012). But most teachers remain unable to create the proposed authentic communicative environment needed to make learners language users, and they try to maintain traditional ways due to many reasons such as classroom reality, their incompetency, inefficient materials, lack of in-service training, and so on. As a consequence, the traditional system has survived in spite of many reforms and innovations (Işık, 2011).

Now, Turkish education system incorporates a four-year primary, a four-year secondary and a four-year high school education for a total of 12 years as compulsory education (Basic National Education Law, 2012). While many foreign languages like English, French, German, Italian, Chinese, Spanish, Japanese, Russian and Arabic are being taught in Turkey, English among them finds the most favored. It is taught in schools from Grade 2 upwards (Ministry of National Education, 2013).

Students are allocated two hours of English classes a week for Grade 2, 3 and 4; three hours of English classes a week for Grade 5 and 6; and four hours of lessons a week for the following grades (Ministry of National Education - Board of Education, 2017; Ministry of National Education - Board of Education, 2018). It may be seen an advantage for a learner since the exposure to language is longer. However, Bayyurt (2012) puts emphasis on the in-class efficiency rather than the number of English classes per week. Some divergences occur in terms of the type and quality of instruction, the number of teaching hours for ELT, the quality of materials and the qualifications of teachers (Kırkgöz, 2005a). For instance, according to results of Education First English Proficiency Index (2017) survey which has ranked countries by their English skills, Turkey is ranked 62<sup>nd</sup> among 80 countries in the world despite the number of English classes and innovations.

#### **2.4. Positive and Negative Aspects of Speaking Tests**

According to Spolsky (2008), “testing has become big business” (p. 297), and it plays a powerful role in education, politics, and society (McNamara and Shohamy, 2008). Aware of this power of tests, policy makers use them to control their educational systems and curricula, to impose new textbooks and new teaching methods, and to prescribe the behavior of administrators, teachers and students (Cheng & Curtis, 2004; Shohamy, Donitsa-Schmidt and Ferman, 1996; Stiggins & Faires-Conklin, 1992). Even though “the testing of speaking is widely regarded as the most challenging of all language tests to prepare, administer, and score” (Madsen, 1983, p. 147), speaking tests have incontrovertible positive aspects.

### **2.4.1. Listening and Speaking**

As S.K. Kitao and K. Kitao (1998) point out “it is difficult to separate the listening skill from the speaking skill. There is an interchange between listening and speaking, and speaking appropriately depends, in part, on comprehending spoken input” (p. 1). Hence, students do not try to only speak but also listen to interlocutor in a speaking test, and it can be sometimes difficult. For example, accent or speech rate of the interlocutor may bring on difficulty in understanding for students. For another example, if content listened to includes unfamiliar linguistic features such as idioms, accent, slang and colloquial language or a local dialect, it also may be difficult for the students. Consequently, students have to cope with these challenges, and then, they have to produce spoken messages in speaking tests properly.

### **2.4.2. Washback**

Washback or backwash refers to “the effect of testing on teaching and learning” (A. Hughes, 2003, p. 1). Washback may be either positive (beneficial) or negative (harmful) (A. Hughes, 2003; Bachman and Palmer, 1996). If a testing procedure supports good teaching and learning practice, it is seen as positive washback (Taylor, 2005). Conversely, undesirable effect of testing on teaching and learning process is known as negative washback (Alderson and Wall, 1993).

For instance, speaking tests elicit strengths and weaknesses of students while students are performing their abilities. Hereby, teachers use tests to determine students’ language learning needs, and it is positive washback. On the other hand, tests are “events, snapshots, brief moments in the process of learning” (Allan, 1999, p. 20), students may not exhibit their full capacity on account of some personal problems like physical, psychological and experiential at that specific moment. As a result, tests may also be inadequate for teachers in eliciting strengths and weaknesses of students (O’Malley & Valdez-Pierce, 1996). Even, they lead teachers to make wrong decisions about students’ language levels, so it is negative washback.

For another example, speaking tests will promote teaching speaking skills. Inasmuch as “what is assessed becomes what is valued, which becomes what is taught” in classes (McEwen, 1995, p. 42). Additionally, testing speaking will be influential in regard to motivating learning. If students achieve in speaking tests, they will feel more confident and use opportunities to speak English both in and out of the classroom. If they fail, they will increase their study time to be successful in speaking skill, which is considered as positive washback. On the contrary, both teachers and students tend to ignore speaking skill owing to difficulty of testing (Allan, 1999; O’Malley & Valdez-Pierce, 1996). As long

as teachers prefer easy way of testing, and they assess grammar and vocabulary knowledge of students rather than speaking skill, the students will not practice speaking skill at all, which is negative washback.

Washback is highly complex notion. Wall (1996) stresses the difficulties in finding the direct causal effects of testing on teaching and learning. Alderson and Wall (1993) indicate that quality of the washback effect and a test do not have to match. Messick (1996) clarifies it, and states that “a poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done in the education system” (p. 242). According to Cheng & Curtis (2004),

Whether the effect of testing is deemed to be positive or negative should also depend on *who* it is that actually conducts the investigation within a particular education context, as well as *where*, the school or university contexts, *when*, the time and duration of using such assessment practices, *why*, the rationale, and *how*, the different approaches used by different participants within the context (p. 8).

However, A. Hughes (2003) suggests the items below in favor of achieving positive washback:

- Test the abilities whose development you want to encourage,
- Sample (tasks) widely and unpredictably,
- Use direct testing (of performance skills),
- Make testing criterion-referenced,
- Base achievement tests on objectives (not textbooks),
- Ensure the test is known and understood by students and teachers,
- Where necessary, provide assistance to teachers,
- Counting the cost (practicality).

### **2.4.3. Reliability**

McNamara (2000) defines reliability as the “consistency of measurement of individuals by a test” (p. 136). In other words, Heaton (1988) explains it in the following way: “if the test is administered to the same candidates on different occasions (with no language practice work between these occasions), then, to the extent that it produces different results, it is not reliable” (p. 162). In this sense, “reliability is a measure of accuracy, consistency, dependability, or fairness of scores resulting from administration of a particular examination” (Henning, 1987, p. 74).

It is argued that speaking test scores do not serve as pure students’ speaking ability like fluency, accuracy, pronunciation, etc. A number of factors almost inevitably enter into testing process of speaking. These factors may contribute to the success/failure of individual students (Harris, 1969; J.D. Brown, 1996) or threaten test reliability. As a result,

unreliable scores bring about unfairly low grades on report cards, wrong placements, or unjustified promotions (Luoma, 2004).

Luoma (2004) and Douglas (2014) illustrate the factors as the language level, gender, the status of the interlocutor, his or her familiarity to the students, the personal characteristics of the interlocutor and students, the nature of the interaction, the sorts of tasks presented to the students, poorly designed test tasks, the questions asked, unclear instructions the opportunities provided to show his or her ability to speak in a foreign language, the strategies the student employs in attempting to complete the test tasks, the criteria used to assess the performance, and the ways in which rating scales are interpreted by an assessor. J.D. Brown (1996, p. 198) also highlights these factors and categorizes them in Figure 2.1.

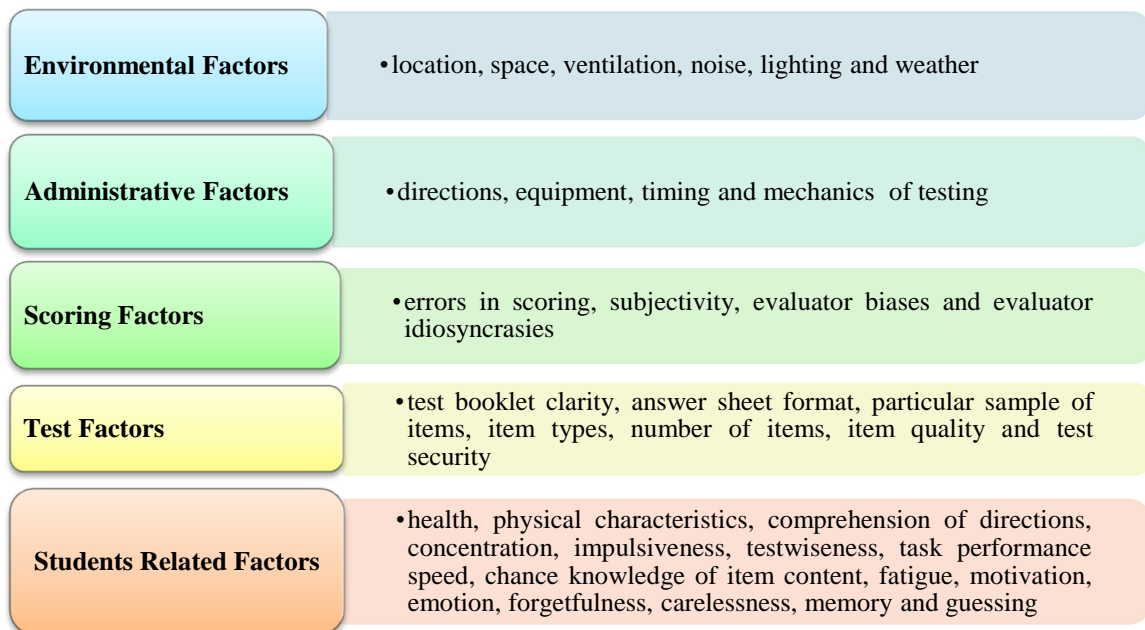


Figure 2.1. Potential factors affecting testing process

*Environmental factors* should be conducive to students' best possible performance: reasonably comfortable seating, comfortable temperature, well-lit, no distracting noises, etc. (Douglas, 2014). Teachers' attitude, students' anxiety, unclear instructions or directions, easy accessible equipments and same timing for each student are likely to have an impact on students' performance in *administrative factors*. Human judgments generally lead to *scoring factors*. Teacher's bias (cultural background, background knowledge, cognitive characteristics, mother tongue, age and gender), motivation, communication style, experience as a rater or an interlocutor, distraction and tiredness during test interfere with reliable scoring (Bachman, 1990; B. Wang, 2010; Luoma, 2004; McNamara, 1996).

In *test factors*, task difficulty has an effect on students' performance in terms of accuracy, fluency, and complexity (Mehnert, 1998; O'Sullivan & Weir, 2002; Skehan,

1996; Skehan & Foster, 1997; Wigglesworth, 1997). For this reason, Douglas (2014) suggests that tests ought to be neither too difficult nor too easy but an appropriate level for students. Furthermore, tests should be of an appropriate length. For example, test tasks for young learners should not be long because they lack the maturity to concentrate for long periods (O’Sullivan, 2013). By all means, a 5-minute speaking test is not as reliable as a 30-minute test. A longer test tends to both minimize the effect of bias and provide a more adequate sample of language use (Bachman, 1990). By the way, there is no clear answer sheet form because speaking tests are more open-ended, and test takers have freedom of choice. So speaking tasks must be specific enough to elicit output within an expected range of performance (H.D. Brown, 2004). Test security can be an issue supposing that some students get a copy of the test previously.

Whereas J.D. Brown (1996) depicts *students-related factors* affecting testing process as in Figure 2.1, O’Sullivan (2000) classifies those under three headings: physical/physiological, psychological and experiential as shown in Figure 2.2.

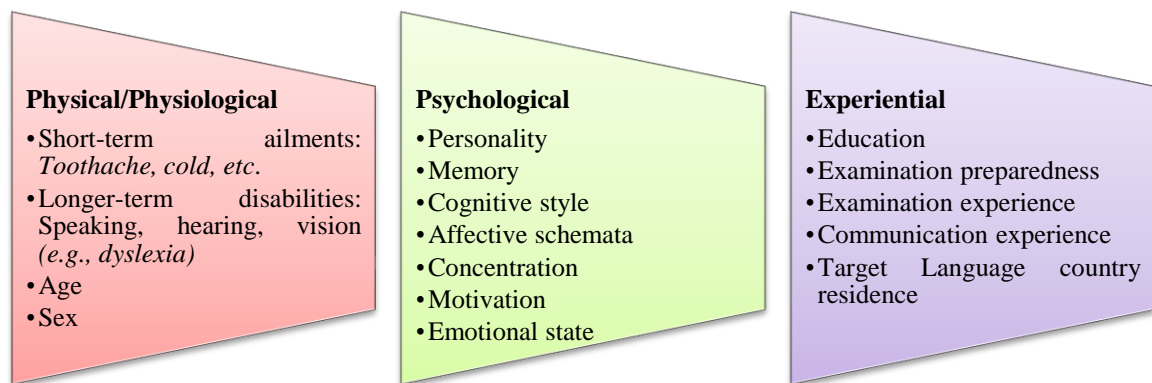


Figure 2.2. Test taker characteristics that affect students’ performance

All of the *physical/physiological characteristics* are likely to have a significant impact on students’ performance. Test designers should be attentive about whether test tasks are appropriate to students’ age and gender. What is more, they should offer an alternative test to students who miss a test due to short-term ailments.

*Psychological characteristics* can also influence students’ performance. For example, Berry (1993) and N. Underhill (1987) discuss students’ personality: on the one hand, N. Underhill (1987) asserts that while extrovert and talkative students are rewarded in speaking tests, unfortunately introvert students can suffer. On the other hand, Berry (1993) reveals that whereas both extrovert and introvert students perform better in a paired test than one to one interview, introverts do not perform as well as extroverts in mixed pairs. By the way, if a student is not at the appropriate level of proficiency in language, they can hardly perform anything.



Much though *experiential characteristics* come from outside of the students, they still have an effect upon students' performance. Henning (1987) sets forth that "tests will tend to be less reliable for persons who have had no prior exposure to the test format and for persons who have developed a high degree of test-wisness" (p. 80). Insofar as, examination experience or test-wisness includes the ability to comprehend easily almost any test directions and performance test tasks, and it may affect the way students deal with the task. Therefore, teachers should clarify test process and exemplify test tasks in advance. Even, they may do test preparation courses.

Consequently, it is important to identify, estimate, control and minimize effects of the factors on the purpose of making tests as accurate as possible and giving students a fair measurement of their abilities (Bachman, 1990; Douglas, 2014). Luoma (2004) recommends teachers to construct the tasks carefully, support themselves with professional development, materials development and rater training with the intent of overcoming some of these problems.

**2.4.3.1. Rater reliability.** The rating is an interaction among the raters, the criteria and the performances in order to produce the scores (Luoma, 2004). In this respect, it is essential to train raters, design effective rating scales and organize standard setting in order to ensure and enhance reliability. Rater reliability is about the consistency among raters. It may be challenging because it involves human raters. As such, this situation creates some variability in the rating process. Teachers may be distracted or tired; they may not fully understand or apply the assessment criteria. N. Underhill (1987) claims that "the more assessors you have for any single test... the more reliable the score will be" (p. 89). Bachman and Palmer (1996) suggest two or three assessors ideally.

There are two types of rater reliability: inter-rater reliability and intra-rater reliability (H.D. Brown, 2004). Fulcher (2014) explains inter-rater reliability as: "test-takers should be able to get the same score irrespective of who is rating their performance" (p. 139). Unless inter-rater reliability can be implemented, it needs to train raters better or redefine criteria more explicitly and detailed (Luoma, 2004). Intra-rater reliability refers to the agreement among scores given by a single rater (H.D. Brown, 2004). A rater has to have his or her own consistency and score same performance in the same way whatever the circumstances.

#### **2.4.4. Validity**

Validity means the appropriateness of a test or any of its component parts as a measure (Henning, 1987). Namely, if an appropriate test measures what it is purported to

measure, the test is agreed as valid. S.K. Kitao and K. Kitao (1998) exemplify a valid test as follows:

If the test purpose is to test ability to communicate in English, then it is valid if it does actually test ability to communicate. If it actually tests knowledge of grammar, then it is not valid test for testing ability to communicate (p. 1).

On the other hand, reliability is a benchmark for validity (Alderson, Clapham and Wall, 1995; Bachman, 1990). A test cannot be valid unless it is reliable. Yet, it is possible for a test to be reliable but not valid.

There are some threats to test validity such as invalid applications of the test, inappropriate selection of content, imperfect cooperation of the examinee (response validity), poor criterion selection (Henning, 1987), imprecise or ambiguous rubrics, unsuitable rating procedures and misinterpretation of test scores.

According to Luoma (2004), in order to enhance validity, test designers firstly need to define and specify the objectives of the test as clearly as possible. Secondly, in the light of these objectives, they construct the test: namely, the tasks and criteria. Then, they administrate the test consistently. Lastly, rating procedures must be fair on behalf of both reliability and validity. Provided that the assessment is fair, it is fit for purpose, and test designers obtain useful, meaningful and reasonable scores. Finally, they implement the construct in parallel with these directions. Common types of validity are content validity, criterion-related validity, construct validity, face validity and response validity (A. Hughes, 2003; Alderson et al., 1995; Bachman, 1990; Heaton, 1988; Henning, 1987; J.D. Brown, 1996).

**2.4.4.1. Content validity.** As A. Hughes (2003) defines, “a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned” (p. 26). In other words, content validity is associated with whether the content of the test is sufficiently representative and comprehensive for validity or not (Henning, 1987). For example, pencil-and-paper tests do not achieve content validity for speaking tests. A speaking test requires students to speak in an authentic context.

It is difficult to identify content of test tasks clearly and unambiguously because there are so many items to assume. For instance, content of a speaking test includes grammatical forms, communication breakdowns and repairs, conversational strategies. What is more, it might include the physical conditions of the room, the seating arrangement, the time of day, the number of people in the room, and the age, sex, and personality characteristics of the teacher and students, etc. (Bachman, 1990).

Content validity is limited in the sense that it focuses on tests, rather than test scores (Bachman, 1990). Whereas a test does not vary across different groups of individuals, the performance of these individuals may vary considerably, and the interpretation of test scores varies for different uses correspondingly (Hambleton, 1978). For example, much as a rating from a speaking test may be a valid indicator of speaking ability, it may not be valid as a measure of teaching ability.

**2.4.4.2. Criterion-related validity.** Criterion-related validity “relates to the degree to which results on the test agree with those provided by some independent assessment of the candidate’s ability” (A. Hughes, 2003, p. 27). For example, there are 50 students for a speaking test, and each student has got 45 minutes during this test. The test has higher validity but it is not practical. Therefore, only 10 minutes are devoted to each student. This time, validity of the test is damaged. In an effort to overcome this problem, randomly selected students are tested for 45 minute as the criterion; the others are tested for 10 minutes. Then the students’ performances on both tests are compared. If the result is a high level of agreement, the 10 minute test may be considered to have criterion-related validity.

There are two kinds of criterion-related validity: concurrent validity and predictive validity. In *concurrent validity*, the test and the criterion are administered at about the same time, and students’ scores on the test are compared with their scores on other language tests. However, it is problem to determine the criterion. Bachman (1990) suggests real life language use as a criterion, yet, he also adds that there is no clear distinction where real life language use begins or ends. In *predictive validity*, a test can predict students’ future performance (A. Hughes, 2003; N. Underhill, 1987). For example, speaking test can predict course success of students or satisfactory job performance.

**2.4.4.3. Construct validity.** A test has construct validity if the right thing is assessed (Luoma, 2004). According to Bachman (1990), it is “the appropriate interpretation of test scores” (p. 255). In this sense, the test scores should give meaningful construct-related information about students’ proficiency. If test tasks include irrelevant variables to the construct or do not include important aspects of the construct, students are unlikely to reflect their true abilities related to the construct (East, 2016). For example, as scoring speaking tests, a teacher evaluates accuracy, fluency, vocabulary, content and pronunciation. Supposing that the teacher evaluated just pronunciation and grammar, construct validity of the test would suffer.

Many linguists argue challenge of construct definition in speaking ability and speaking tests. Luoma (2004) asserts three approaches to define the speaking construct for assessment. *The linguistic approach* focuses on language forms: vocabulary, grammar,

pronunciation and fluency, etc. *The communicative approach* focuses on how well the students can use the skills and strategies that test activities require. *The task-based approach* focuses on the students' ability to deal with test tasks.

**2.4.4.4. Face validity.** A test has face validity if it looks right, and appears to measure the knowledge or abilities it claims to measure (Mousavi, 2002). It is concerned with tests' "surface credibility or public acceptability" (Ingram, 1977, p. 18). According to H. D. Brown (2004),

face validity will likely be high if learners encounter

- a well-constructed, expected format with familiar tasks,
- a test that is clearly doable within the allotted time limit,
- items that are clear and uncomplicated,
- directions that are crystal clear,
- tasks that relate to their course work (content validity), and
- a difficulty level that presents a reasonable challenge (p. 27).

Face validity is commonly regarded as unscientific and irrelevant (Stevenson, 1985) since this validity is not tested by an expert or a teacher, and it includes judgments of people like students or administrators. However, many advocates of CLT such as B.J. Carroll (1980), Morrow (1986) attach importance to the appearance of the test and plead that the appearance of the test should be from the real world. In order to enhance face validity, tests designer had better show the test to other professionals and get their opinion (Heaton, 1988; N. Underhill, 1987).

**2.4.4.5. Response validity.** According to Henning (1987), "response validity is related to whether students respond in the expected manner or not" (p. 96). Students may not reveal their actual ability by the virtue of some reasons such as unclear instructions, unfamiliar test format. In that case, the test may lack response validity.

## **2.4.5. Practicality**

Practicality is concerned with test implementation. H.D. Brown (1994) suggests that "a test ought to be practical – within the means of financial limitations, time constraints, ease of administration, and scoring and interpretation" (p. 253). Bachman & Palmer (1996) construe practicality as "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities" (p. 36). However, it is nearly impossible to assess speaking skill in a practical way.

Dimensions of impracticality generally focus on both administration problems and workload. Improper administration can cause students to perform less effectively. For example, increased noise during a test can result in reduced scores (Haines, Stansfeld, Head and Job, 2002; Powers et al., 2002) because it affects concentration. Speaking tests

are time-consuming in that they are usually carried out “during a face-to-face interaction, in real time, between an interlocutor and a candidate” (Luoma, 2004, p. ix). “Even if each student speaks for only a few minutes, this becomes a huge job,” especially in large numbers of students (S.K. Kitao and K. Kitao, 1998, p. 1). Besides, it is difficult to supply the necessary number of interlocutors and assessors, and their training.

Recording speaking performance is fundamental. It is generally seen as technical difficulty in the point of poor quality of sounds (N. Underhill, 1987; Weir, 1990). During the test, “the assessor has to make instantaneous judgment” (Luoma, 2004, p. ix) unless he or she records speaking test. In this sense, recording increases test reliability (Jones and Spolsky, 1975). All the same, it is time-consuming since the assessors have to listen to each performance second by second after the speaking test. In addition, teachers can keep records for future reference in case of appeal, not for scoring students later. It is still difficult with regards to storing recordings.

#### **2.4.6. Objectivity**

Objectivity is difficult in speaking tests, especially in the large scale testing (Spolsky, 1990). Harris (1969) notes that any assessor cannot maintain exactly the same scoring standards throughout a large number of students. Unfortunately, it is likely unfeasible to avoid subjectivity in assessment (Weir, 1990; Alderson et al., 1995). Notwithstanding, this problem can be still minimized by standardization.

#### **2.4.7. Standardization**

According to Cohen and Wollack (2006), “tests are standardized when the directions, conditions of administration, and scoring are clearly defined and fixed for all examinees, administrations, and forms” (p. 358). Well-considered standards regardless of their troubles are better than no standards at all (Hambleton, 1978; Popham, 1978; Scriven, 1978). Insofar as, no standards presumably mean that tests are being done unsystematically, and perhaps, unfairly. Standardization of the tests makes them more valid and reliable because all students are expected to do the same thing under the same conditions (Allan, 1999), and hence, teachers decide more confidently in the direction of the test purpose as well.

#### **2.4.8. Authenticity**

Test authenticity is “the degree of correspondence of the characteristics of a given language test to the features of a target language use task” (Bachman and Palmer, 1996, p. 23). As such, real life performance is a criterion for authenticity. The real life approach has been dominant in the area of testing speaking skill in recent years (Bachman, 1990). This

approach dissents artificial language usage in the course of speaking test and encourages language usage in real or authentic situations. Authenticity appreciates quality of the language assessment. In this regard, according to H.D. Brown (2004), authenticity in a test may be presented as follows:

- The language in the test is as natural as possible,
- Items are contextualized rather than isolated,
- Topics are meaningful (relevant, interesting) for the learner,
- Some thematic organization to items is provided, such as through a story line or episode,
- Tasks represent, or closely approximate, real-world tasks (p. 28).

Authentic tasks open the way to connect directly with the real world and engage students to produce a meaningful outcome. Accordingly, students are encouraged to use the target language by responding, reacting, and interacting no matter whether they speak accurately or not. If teachers want students to perform their speaking skills in real world, teachers also need to create assessment opportunities that allow them to use the type of language they are likely to encounter beyond the classroom (East, 2008).

The following is a case in point: choosing a foreign chef who will work in a hotel kitchen in Paris.

The test can take place in a model kitchen, complete with everything necessary for food preparation, including the food. The test taker will have to interact with French-speaking colleagues to discuss the day's menu, select appropriate ingredients, and prepare an actual dish. The performance will be observed systematically for accurate use of vocabulary and syntax, appropriate professional manner, and correct use of recipe and procedural genres, and given a score reflecting the observers' (or raters') assessment of the chef's abilities in French (Douglas, 2014, p. 24).

Spolsky (1985) points out "lack of authenticity in the material used in a test raises issues about the generalizability of results" (p. 39). In other words, students' performance on a test cannot be generalized in other contexts unless the test is authentic. Poor speaking tests evaluate memorization skills of students. On the one hand, memorization possibly gives lower level students a back and contributes to their fluency. On the other hand, well-memorized responses of students put an end to spontaneity during speaking tests. However, both fluency and spontaneity bring up authenticity of speaking tests.

#### **2.4.9. Interactiveness**

Interactiveness is identified with "the extent and type of involvement of the test taker's individual characteristics (language ability, topical knowledge and affective schemata) in accomplishing a test task" (Bachman & Palmer, 1996, p. 25). Speaking tasks in a test had better not provoke negative feelings of students. Students need positive interaction with the tasks. For example, they can select topic of the task or form the structure of their interaction; thereby, they can demonstrate their full capacity in terms of their language ability.

#### **2.4.10. Test Anxiety**

Speaking tests provide “the greatest opportunity for test takers to exhibit their best performance” (Bachman, 1990, p. 156). In this case, this opportunity can bring about stress, pressure and anxiety as well. Test anxiety can be a barrier blocking a student on performing their full capacity (Heaton, 1988; Huerta-Macias, 1995). Furthermore, teachers can also suffer from test anxiety on the grounds that poor performance of the students draw on poor results of test which make teachers feel guilty and embarrassed (Alderson & Wall, 1993).

With the aim of overcoming this anxiety, speaking tasks should not be lost in the shuffle during class time. The testing procedure had better reflect the skills and abilities that are taught in the course. When students participate in necessary speaking activities in classes, their readiness level will be high, and they will perform at their best. Otherwise, students are unlikely to perform well on tasks they have not previously met. If they feel familiar with speaking tasks from English classes, their anxiety level may go down. In the contrary case, for example, students have very communicative and interactive activities in classroom but the test contains only multiple-choice tasks away from communication. By all means, anxiety levels will rise, and also students will ignore class activities.

### **2.5. Test Specifications**

Fulcher and Davidson (2007) define test specifications as “generative explanatory documents for the creation of test tasks” (p. 52). Fulcher (2014) takes this definition a step further and points out that “test specifications are dynamic, evolving, documents that should be related to the process of test design, piloting and revision” (p. 116). According to Alderson et al. (1995), task specifications should illustrate the test’s purpose (placement, achievement, proficiency, etc.), description of the test taker (age, sex, level, cultural background, etc.), test level, definition of construct, description of suitable language course or textbook, number of tasks, time for each task, weighting for each task, target language situation, text-types (written/spoken, topic, etc.), text length, language skills to be tested (discrete/integrated, macro/micro skills, etc.), language elements to be tested (structures, functions, notions, etc.), test tasks (authentic, live/recorded, etc), test methods (role play, interview, discussion, etc.), rubrics, criteria for scoring, descriptions of typical performance at each level, description of what students at each level can do in the real world, sample papers, and samples of students’ performance on task.

In conclusion, test specifications characterize purpose, level, construct, setting, time arrangement, administration plans, instructions, materials, linguistic features, and rating criteria of the test (Luoma, 2004). They are useful because they make test development

more planned and avoid problems during test administration or scoring. Furthermore, test specifications give users a clear view of the theoretical underpinnings of the assessment (Luoma, 2004).

## **2.6. Techniques for Testing Speaking**

Bachman and Palmer (1996, p. 44) define test task “as an activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular speaking situation”. Appropriate tests are fundamentally selected or designed on the basis of test specifications. Even so, speaking tests ought to provide opportunities for students so as to show what they know and how they interact (Luoma, 2004). Moreover, they should be comprehensible and motivating in order to elicit appropriate output.

### **2.6.1. Discrete or Integrated Tests?**

Even if discrete tests measure just one skill at one time, integrated tests mean combination of skills (reading, listening, writing, and speaking) in evaluation. Literally, it is impossible to test speaking skill purely apart from other skills (especially listening or reading) because students take input in written or spoken form (Douglas, 1997). Yet, separate scoring is possible; as such teachers are just focused on speaking skills in the course of assessment. The advantages of integrated assessment are authenticity and the use of context (Luoma, 2004; Weir, 1990). Students first comprehend input in a context, and then production comes in compliance with the context. As a consequence, input affects production in terms of fluency, contextualization and organization of content. However, integrated tests are challenging to produce, take, score, and interpret (Weir, 1990). A. Brown, McNamara, Iwashita and O’Hagan (2001) also determine that students perform worse in the integrated than discrete tests because students have to deal with two or more skills.

### **2.6.2. Individual or Paired/Group Tests?**

Individual tests include interaction between a student and a teacher (interlocutor). They are suitable for flexibility in that teachers can adapt their test tasks in conformity with students’ level, interest, etc. However, teachers dominate test tasks and have huge power over students (Luoma, 2004; Ross & Berwick, 1992; Taylor, 2001; Young & Milanovic, 1992). This power influences students’ performance and authenticity of the test negatively (Együd and Glover, 2001). What is more, Weir (2005) states that individual tests are “time-consuming and expensive to administer when conducted on a large scale” (p. 156).

In paired or group tests, students are asked to interact with peers in a variety of tasks. Each student needs to have equal opportunity in these test tasks, yet a student may



dominate the test by giving minimum chance to the other, or students may not complete the test due to mismatch (Weir, 1993; Foot, 1999). Therefore, teachers initially need to be picky while they are forming pairs or groups. During test, the teachers purely observe them “rather than taking part in the interaction directly” (Luoma, 2004, p. 36). Weir (2005) pleads paired and group assessments instead of individual tests, and claims that these tests encourage real communication, cooperation and participation. According to May (2009), they also help to develop the students’ interactional competence and conversation management skills. In addition, paired and group assessments provoke less anxiety and more confidence in students (A. Hughes, 2003; Fulcher, 1996; Ockey, 2001) so they perform better compared to the individual tests (L. Brooks, 2009).

On the one hand, paired or group assessments are practical in terms of timing, as teachers assess two or more students simultaneously rather than assessing one by one (Ducasse & Brown, 2009; East, 2016; Galaczi, 2010; Ockey, 2001; Swain, 2001). On the other hand, administrative problems can occur owing to the sizes of groups or the mixture of ability levels (Reves, 1991). What is more, teachers have to cope with scoring challenges; as such teachers need to separate each student’s performance as scoring because each student’s performance are interdependent much though they are assessed together (L. Brooks, 2009; Swain, 2001).

### **2.6.3. Direct or Indirect Tests?**

While there is face-to-face interaction with a human in direct tests, there is no face-to-face interaction in indirect tests. In indirect tests, students perform their speaking skills by recording their talk with regard to input, and then teachers listen to the recordings, finally they score the students’ performance. The lack of reciprocity makes indirect tests artificial (Luoma, 2004; A. Hughes, 2003). In the meantime, indirect tests become less stressful since students do not see anyone. Moreover, indirect tests offer standardized input that is designed in advance rather than unpredictable and natural talks in direct tests so indirect tests are inflexible from the point of context, response adaptation and creativity compared to direct tests. According to Luoma (2004), standardized input provides convenience in comparability and scoring. A. Hughes (2003) claims indirect tests can be economical in that large numbers of students can be tested at the same time. However, scoring takes time seeing that teachers have to listen to the recordings again and again.

#### 2.6.4. Types of Test Tasks for Speaking Assessment

According to Bygate (1987), types of test tasks in speaking skills are informational and interactional tasks. Informational tasks are presented in Figure 2.3, and interactional tasks are clearly visible in Figure 2.4.

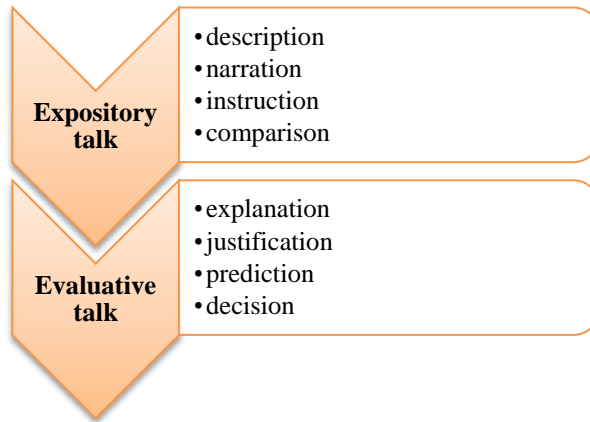


Figure 2.3. Informational tasks

In *description tasks*, students are asked to describe something like a picture, a person, a place, an event, etc. In *narrative tasks*, students are asked to narrate a personal story or picture sequences. Picture sequences are generally preferred because students may feel ashamed in their personal stories. To do well on the tasks, students should “show their control of the essential features of narratives: setting the scene, identifying the characters and referring to them consistently, identifying the main events, and telling them in a coherent sequence” (Luoma, 2004, p. 144). In *instruction tasks*, teachers give directions and instructions. Students get the meaning and show it. The tasks are based on understanding. In *comparison tasks*, students can compare objects, pictures, people, graphs, events, notions or concepts. They analyze and discuss the similarities and differences or the advantages and disadvantages.

In *explanation tasks*, students are asked to explain a process, an event, and contents of a graph. As explaining; students at first set the scene, next identify parts of the information, and lastly explain the significance of these parts so as to clarify topic and reasons for explanation (Luoma, 2004). *Prediction tasks* generally used with explanation tasks feature the uncertainty, and this makes way for negotiation. *Decision and justification tasks* involve discussion. In the process of discussion, students first express their opinions about discussion topic; next, they try to justify their own views and explain the reasons for or against the topic; finally, they make a decision.

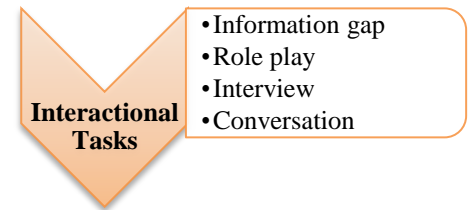


Figure 2.4. Interactional tasks

In *information gap tasks*, “... one of the students has information, the other lacks and vice versa. The students are expected to reach a conclusion by exchanging information in a limited period of time” (Coombe, Folse and Hubley, 2007, pp. 119-122). In this context, the tasks are authentic and purposeful because students genuinely have to communicate to fill an information gap. Role play, interview and conversation tasks will be discussed in the following paragraphs.

According to Luoma (2004), types of tasks in speaking assessment are open-ended, semi-structured and structured tasks as shown Figure 2.5.

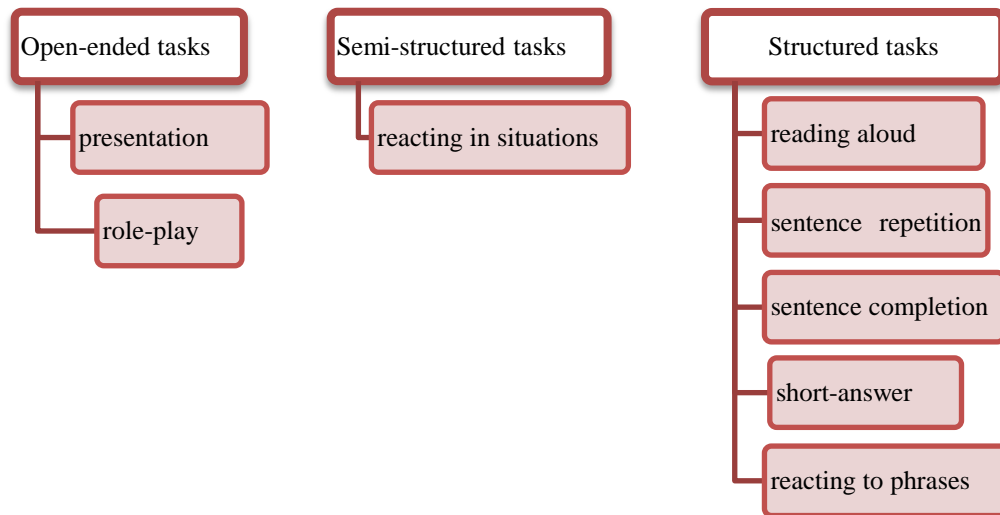


Figure 2.5. Types of tasks in speaking assessment

*Open-ended tasks* give students more freedom as speaking and test more productive skills of students. However, “open-ended techniques can be rather difficult to score” (Madsen, 1983, p. 162) because the tasks assess the unpredictable and creative elements of speaking. In *structured tasks*, students’ output is usually short and predetermined. Even teachers can form a list of acceptable responses in order to compare students’ performance and score easily. The structured tasks typically used to test students’ linguistic competence especially pronunciation and grammar. Examples of open-ended and structured tasks will be discussed in the following paragraphs. *Semi-structured tasks* are more flexible than structured ones even if students’ output is controlled. *Reacting in situations* as a semi-structured task is a really difficult task for students in that they have to adapt themselves to different roles. Students firstly read or hear a social situation, secondly they adapt themselves this situation, and lastly they speak according to the accepted situation.

According to H.D. Brown (2004), types of assessment tasks in speaking skill are imitative, intensive, responsive, interactive and extensive assessment tasks (See Figure 2.6).

Imitative	• word/phrase repetition, sentence repetition
Intensive	• directed response tasks, read aloud, sentence/dialogue completion, oral questionnaires, picture-cued tasks, translation
Responsive	• question and answer, giving instructions and directions, paraphrasing
Interactive	• interview, role play, discussions and conversations, games
Extensive	• speeches, oral presentations, picture-cued story-telling, retelling a story/news event, translation

Figure 2.6. Types of assessment tasks in speaking skill

**2.6.4.1. Imitative tasks.** Students reproduce a word, phrase or a sentence accurately in imitative assessment tasks. Thus, there is no need for negotiation or interaction. Students simply retain teachers' message to imitate. In *sentence repetition tasks*, sentences become longer and more complex in the process so students' memory is crucial. Teachers generally score students' pronunciation skills such as stress and intonation.

**2.6.4.2. Intensive tasks.** Intensive assessment tasks are also valued as *limited response tasks* (Madsen, 1983) or *mechanical tasks* (N. Underhill, 1987). Students prove their grammatical, phrasal, lexical phonological or semantic competence. Notwithstanding students productions are controlled, thus scoring is easy because teachers can make comparisons among students' performance easily.

In *directed response tasks*, teachers direct specified structure or sentence transformation. The following is an example of the direction: tell me you enjoy rhythm and blues, or tell me he can speak Japanese. Hence, the tasks need imitation or modification. *Read-aloud tasks* are practical. Teachers select a reading passage and apply it for all students. Read-aloud tasks checks students' pronunciation skills like rhythm, stress and intonation rather than appropriateness of response or fluency so they are inauthentic. One more drawback of these tasks is the fact that reading skills may interfere while just speaking skills are supposed to be evaluated (A. Hughes, 2003). In *sentence/dialogue completion tasks*, students at first read context which has omitted parts previously, next figure out blank parts, and finally complete the context in a meaningful way. The tasks avoid aural misunderstanding because input is in the form of written. It is an advantage for students since they have more time to anticipate an answer.

In *form-filling* or *oral questionnaire tasks*, students take a form or a questionnaire and answer the questions orally. *Picture-cued tasks* in intensive level are designed to elicit

a word or a phrase. They need description and story-telling in extensive level. Pictures should be “clear and unambiguous and free from cultural or educational bias” (Weir, 2005, p. 148) thus, students can demonstrate their best performance. Besides, they should not be too complex or easy to interpret (Luoma, 2004). Appropriate diagrams, graphics, charts, cartoons, newspaper or maps apart from pictures can also be utilized as task materials. In *translation tasks*, teachers say a word, a phrase or a sentence, and ask students to translate it. It is optional to give thinking time according to task level. The tasks are meaningful especially for interpreter education.

**2.6.4.3. Responsive tasks.** Responsive assessment tasks include comprehension and limited interaction. In *question and answer, short-answer* or *reacting to phrases tasks*, teachers ask questions or give phrases, and students comprehend them, and then give short response. The questions should be purposeful in order to test students’ discourse competence. In *giving instructions and directions tasks*, teachers pose a problem such as “how do I make dinner for thanksgiving? or how do I get to...?”. Students respond with appropriate instruction or direction. The tasks should not evaluate whether students know general knowledge like making dinner or not. They need to focus on speaking skills of the students. In *paraphrasing tasks*, students hear or read five or six sentences and paraphrase them orally.

**2.6.4.4. Interactive tasks.** Interactive assessment tasks include more complex interaction. Students exchange meaning and maintain communication. *Role-play tasks* put students in varied roles which help students to demonstrate their interactive skills. The students structure their talk according to the role-play situation which is explained at the beginning of test. *Simulations* like role-plays try to reproduce real situations. They “involve acting out an imagined communication situation” (Luoma, 2004, p. 153). The main disadvantage of the tasks is the fact that “acting ability can influence the performance unfairly” (Luoma, 2004, p. 153). *Discussion, conversations and games* offer authenticity and spontaneity. Students concentrate on the tasks, not language items. Yet, scoring is challenging issue.

*Interviews* are “the most common format for the testing” (A. Hughes, 2003, p. 119). Students answer the questions whereas the interlocutor asks questions. They are used to “gather information regarding an individual’s experiences and knowledge; his or her opinions, beliefs, and feelings” (Best and Kahn, 1998, p. 255). According to Madsen (1983), “the oral interview can provide a genuine sense of communication” (p. 162). However, Jones and Spolsky (1975) view interview as unnatural tasks. Students know their mark depend on the task, and their talk are for the purposes of the test after all.

The sorts of interviews are free and controlled interviews (Weir, 2005). In the *controlled interview*, predetermined procedures are applied. Interlocutor's questions, instructions and prompts are standardized in the direction of specified topic and structures (Weir, 2005). As such teachers can make comparison among students easily (Fulcher, 2014) and score consistently. However, it does not belong to the spirit of neutral interaction. The *free "interviews* are like extended conversations" (p. 153) and there is no predetermined procedure (Weir, 2005). It is not teacher directed interaction in contrast to controlled interview. The nature of interaction is flexible in that the pace, scope and level of the interaction can be modified easily in these tasks. Students are able to affect the context, "take the initiative, change the direction of the interaction and introduce new topics" (p. 153), take on responsibility for keeping the interaction (Weir, 2005). They have an opportunity to show their capability creatively.

**2.6.4.5. Extensive tasks.** Extensive assessment tasks are monologue. The tasks are based on minimal interaction. Teachers generally do not interfere, but they should be attentive while assigning topics because the topics may affect students' performance (Weir, 2005). In *Oral presentations*, students prepare a report, a paper, an essay on a topic, and present it. Coombe et al. (2007, p. 122) emphasize that "oral presentations are not just concerned with language accuracy and fluency. They also include aspects of delivery such as body language, facial expression, eye contact, and gestures". In *picture-cued story-telling tasks*, series of pictures are particularly useful for story-telling while single pictures are viewed for descriptions in extensive level (A. Hughes, 2003). These tasks minimize the amount of oral input during test.

In *retelling a story/news event tasks*, students hear or read a story or news event, and then they retell it. In *translation tasks*, longer texts are presented for students to translate them in extensive level. The tasks control knowledge of structures and vocabulary, content, and comprehension. They test production in that students convey the meaning of the text as well. In conclusion, translation tasks need highly specialized skill for students.

## 2.7. Test Administration

Procedures for test administration are suggested by different scholars (Bachman & Palmer, 1996; Douglas, 2014; Fulcher, 2014; Fulcher & Davidson, 2007; Luoma, 2004) as follows:

- The procedures should be planned carefully and organized well. Thus, the assessment process flows smoothly.

- Teachers should design appropriate test tasks and assessment criteria in advance.
- Teachers should schedule test location and time clearly, later they should announce it to the students.
- Testing environment should be laid out well beforehand. Test rooms such as interview rooms, waiting rooms, and language laboratory should be clean and well aired. Ideally, test rooms should have comfortable furniture, adequate work space, appropriate heating or cooling, good lighting, lack of distractions like noise, and be reasonably quiet.
- Students should be aware of timing: when the test will start, how long it will take, and whether the students or teachers need break time, etc. Moreover, each student should have sufficient time to show the best of their speaking ability during test.
- Teachers should create peaceful and supportive atmosphere for the students. Namely, they should not arouse a negative feeling in students.
- Teachers should provide written or oral instruction. Students should know how they will prepare for the test, what they will do during the test, what criteria will be applied, how they will be assessed.
- Any materials (picture, newspaper, realia, diagram, map, etc.) and equipment (microphones, speakers, recorders, computer, etc.) should be checked and ready for use.
- An adequate number of teachers are assigned as interlocutors, assessors or administrators.
- Teachers should be well-supported with professional development, materials development, and rater training.
- As soon as the test is over, all test components must be collected.
- Teachers declare students' marks after scoring.

### **2.8. Rater and Interlocutor Training**

In speaking tests, “one teacher, the interlocutor, interacts with the student or students being tested. The other teacher, the assessor, focuses on writing scores and making notes” (Coombe et al., 2007, pp. 115-116). N. Underhill (1987) points out “an interlocutor is not an assessor” (p. 7). Interlocutors engage students in conversation via the input associated with the tasks (Fulcher & Davidson, 2007). They should not be seen making notes about a students' performance (A. Hughes, 2003) because it can distract the student and create unnecessary anxiety (Nagata, 1995). Assessors, in other words scorers

or raters, “judge performances in productive tests of speaking and writing, using an agreed rating procedure and criteria in so doing” (McNamara, 2000, p. 136).

Wilkinson (1968) advocates that the training of interlocutors and raters is essential for standardization of their roles. Besides, this training helps teachers to develop common interpretation of the rating scales. The obvious advantage is that the training enhances reliability, validity and fairness, and objectivity (Douglas, 2014; Fulcher, 2014; Fulcher & Davidson, 2007; J.D. Brown, 1996; Luoma, 2004). Another advantage of the training is that teachers feel more confidence and competence; however, it cannot guarantee that all teachers will give marks as they are supposed to since each rater has idiosyncratic perception, beliefs and preference.

The advice of the British Council’s *VOTE* (1983) to interlocutors is as follows:

- Do not correct the test taker when they make mistakes.
- Do not speak so quickly that the test taker has difficulty understanding you.
- Do not whisper, cover your mouth or mumble.
- Do not speak too much.
- Do not be condescending (e.g. following an error in speaking, do not say “it is a bit difficult, is not it, speaking English”).
- Do not be offensive (e.g. make negative comments about the test taker’s culture, etc.).
- Maintain eye contact with the test taker.
- Do not engage in other activities (e.g. reading the assessment criteria, or candidate forms) during the test.

Rater training procedures are defined as follows:

- ❖ Introduce the test, the scale, criteria and levels of the scale.
- ❖ Analyze samples that have been previously rated by experienced raters, and discuss how they rate and apply criteria.
- ❖ Practise rating, compare the ratings with samples.
- ❖ Rate more, discuss reasons for the scores and get the consensus score (Bachman & Palmer, 1996; Luoma, 2004).

## **2.9. The Common European Framework of Reference for Languages (CEFR)**

The CEFR is a significant source for language education. Its purpose is to provide a common basis for the elaboration of language curricula, syllabuses, course materials, teaching/learning process, assessment, etc. across Europe (Council of Europe, 2001). Accordingly, the CEFR is intended to:



- “promote and facilitate co-operation among educational institutions in different countries;
- provide a sound basis for the mutual recognition of language qualifications;
- assist learners, teachers, course designers, examining bodies and educational administrators to situate and co-ordinate their efforts” (Council of Europe, 2001, pp. 5-6).

In language assessment, the CEFR defines the levels of language proficiency to facilitate comparisons between different systems of qualifications, and it also proposes useful descriptors at each level. These levels are basic user (A1 & A2), independent user (B1 & B2) and proficient user (C1 & C2) (Council of Europe, 2001). The descriptors for each level change with respect to communicative activities categorized in the CEFR. In speaking skills, communicative activities are divided into spoken production and spoken interaction as can be seen in Figure 2.7 (Council of Europe, 2018).

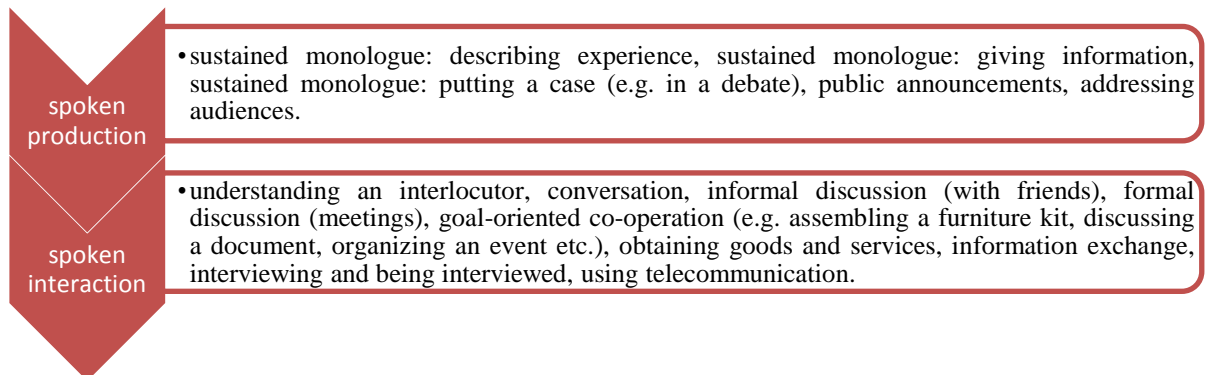


Figure 2.7. Activities of spoken production and interaction

In lieu of detailed descriptors for all activities in the CEFR, the descriptors of overall spoken production and interaction at each of six levels are provided in Table 2.2 and Table 2.3 (Council of Europe, 2018). The descriptors provide teachers convenient criteria as they are assessing students’ abilities.

Table 2.2

*The Descriptors for Overall Spoken Production*

C2	Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.
C1	Can give clear, detailed descriptions and presentations on complex subjects, integrating sub themes, developing particular points and rounding off with an appropriate conclusion. Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.
B2	Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.
B1	Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.
A2	Can give a simple description or presentation of people, living or working conditions, daily

	routines, likes/dislikes etc. as a short series of simple phrases and sentences linked into a list.
A1	Can produce simple mainly isolated phrases about people and places.
Pre-A1	Can produce short phrases about themselves, giving basic personal information (e.g. name, address, family, nationality).

Table 2.3

*The Descriptors for Overall Spoken Interaction*

C2	Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.
C1	Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language. Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances.
B2	Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with speakers of the target language quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments. Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc.
B1	Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. Can enter unprepared into conversation of familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events). Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations.
A2	Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of his/her own accord.
A1	Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech, rephrasing and repair. Can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.
Pre-A1	Can ask and answer questions about him/herself and daily routines, using short, formulaic expressions and relying on gestures to reinforce the information.

**2.10. Rating Scales**

Rating scale is an interchangeable term with *scoring rubric* or *proficiency scale* (Fulcher, 2014; Fulcher & Davidson, 2007). N. Underhill (1987) defines rating scale as “a series of short descriptions of different levels of language ability” (p. 98). According to Fulcher (2014), “the purpose of the rating scale is to guide the rating process” (p. 89). Therefore, a rating scale is directly connected with the purpose of the test, the construct definition of the test, the tasks and the criteria (Luoma, 2004). Teachers use rating scales to score speaking performances of the students consistently. It enhances reliability of the

speaking tests. What is more, it gives students a bunch of expectations about what will be assessed.

Rating scales comprise a set of levels upon which students' performances are judged (Davies et al., 1999; Wigglesworth & Frost, 2017). Each level from excellent to poor is defined by a descriptor, and it commonly characterizes linguistic features (vocabulary, syntax, fluency, etc.), tasks and functions students can perform (Davies et al., 1999; Fulcher, 2010). According to Luoma (2004), the number of levels is also important in a rating scale which often has four to six. She sets forth "the more levels there are, the more specific the feedback will be" (p. 80), and the lower levels there are, the more consistent raters' decisions will be. Types of rating scales are holistic and analytic scales (A. Hughes, 2003; Council of Europe, 2001; Fulcher, 2014; Luoma, 2004; Madsen, 1983). While holistic scales provide valuable information for an overall categorization of speaking ability, analytic scales provide more diagnostic information.

*Holistic scales* or *global scales* (see Appendix A) are on the basis of an overall impression. Raters match students' performance with one of a range of descriptions on scale. Teachers who have enough experience and specialized training tend to select holistic scale (Madsen, 1983; N. Underhill, 1987). Unlike analytic scales, raters do not have to score each criterion separately so holistic scales are practical, simple and speed (H.D. Brown, 2004; Luoma, 2004; O'Sullivan, 2008; Weir, 2005). But it is not easy to interpret students' scores because each rater has his own criteria in his mind. Furthermore, it does not provide useful feedback for students in order to improve their speaking skills. To sum up, although holistic scales can bring some advantages to teachers in assessing students' speaking skills, they offer some disadvantages to students in their autonomous process of learning speaking skill (Tuan, 2012).

*Analytic scales* (see Appendix B) have been found more reliable than holistic scales even though holistic scales are acceptable. Analytic scales include a number of criteria such as accuracy, fluency, pronunciation, etc., and each criterion has descriptors at the different levels of the scale (Luoma, 2004). Raters need to decide how each criterion will be weighted because some criteria may be weighted more heavily, or vice versa. Compared to holistic scales, analytic scales are particularly useful for inexperienced raters to train and standardize them (Weir, 2005).

While using analytic scales, raters have to score each criterion separately. On the one hand, this detailed guidance is clear advantage of analytic scales (H.D. Brown, 2004; Luoma, 2004). Hence, the scales help raters to diagnose strong and weak aspects of students' performances and give feedback to students (Fulcher & Davidson, 2007; Luoma,

2004; Weir, 2005). In addition, students can follow their progress over time in some or all dimensions. On the other hand, analytic scales are so detailed that raters have to check, consider, and score each criterion; therefore, it is time consuming. What is more, concentration on the different aspects may divert attention from the overall effect of the speech (A. Hughes, 2003), and raters can oversimplify the components of the performance.

### **2.11. Research Studies on Students' and Teachers' Attitudes towards Speaking Tests**

Şallı-Çopur (2002) examined students' oral performance in speaking test methods (oral presentation, role play, individual and paired interviews) and their attitudes towards being tested through these test methods. The participants were 25 first year students of the Foreign Language Education Department at Middle East Technical University. The analysis of data collected through the speaking tests indicated that the students were positive towards all four methods. Moreover, they also showed a positive attitude towards being recorded or videotaped during the speaking tests. However, they expressed preference towards being tested with individual interview rather than with paired interview.

Restrepo et al. (2003) searched teachers' understanding of oral assessment so as to establish similar oral evaluation criteria for teachers, students, and the institution in Columbia. 30 teachers were interviewed on their beliefs and practices on oral assessment. The analysis of the data revealed that most teachers had a weak knowledge of assessment. Teachers were well aware of what and how to assess. However, they did not plan testing process in detailed, and they evaluated the students unsystematically for summative purposes. Moreover, very few teachers gave reasons of assessment as a process through which teaching methodology and learning could be improved. As a result, the study recommended similar criteria for assessment, which helps teachers to assess their students more systematically and objectively. The researchers also put an emphasis on the need of in-service training courses in the area of assessment for the teachers.

Sook (2003) conducted a research associated with the types of speaking assessment tasks used by Korean Junior Secondary School English Teachers. She also investigated Korean teachers' perceptions of the practical constraints in Korean EFL classrooms which affected speaking assessment. The participants were 10 Korean Junior Secondary School English teachers. Qualitative methods were adopted. All the participants responded to a questionnaire which asked for opinions on speaking assessment. Some of them also participated in an interview with the researcher to clarify their perceptions of speaking assessment. The conclusion was reached that the speaking assessment tasks (a) gave the students less psychological burden, (b) were time-saving and designed for the convenience

of construction and administration, (c) did not demand the teacher to take the role of an interviewer. What is more, the study found out that the participants were not concerned with the validity and reliability of their assessments. They also lacked training in conducting speaking assessment. Ultimately, the study suggested that Korean teachers needed to make themselves conscious and persistent efforts in order to introduce more communicative speaking assessment in spite of practical difficulties.

Güllüoğlu (2004) studied on whether speaking was given the necessary emphasis at Gazi University Preparatory School of English. The participants were 127 students and 73 instructors at prep school. Two types of questionnaires were conducted in this study: the first one was administrated before the speaking test, and the second one was applied after the speaking test. While the first questionnaires aimed to find out how much emphasis was given to speaking skills in class, the second questionnaires were to reveal the attitudes of both instructors and students towards speaking skills conducted at prep school. The analysis and interpretation of the first questionnaires yielded that grammar, reading and vocabulary were the skills that were mostly emphasized though the students wanted speaking skills to be given the most weight in class. The majority of the students did not feel confident while speaking in English, so they indicated their unwillingness to take speaking tests. The analysis of the second questionnaires illustrated that both the instructors and the students adopted positive attitudes towards the speaking test which was applied after the first questionnaire. They realized that the speaking test they took was not as difficult as they expected. They became aware of the fact that they could succeed in these tests if speaking were given more weight in the curriculum and class. The study indicated the necessity and usefulness of a speaking course and speaking tests at Gazi University Prep School.

Majid, Eng & Samad (2007) studied with eight English language teachers. The purpose of this study was to analyze the effects of the school-based oral English test (OET) on teacher perceptions and practices in three MARA Junior Science Colleges in Malaysia. From the data obtained through the interviews and observations, a few teachers believed that the OET was a fair test of speaking. Rehearsing and memorization were preparation stages for the test, and few teachers advocated the use of memorized text in oral testing. The OET affected the teachers: they became more concerned with how best to assess the students' spoken language, and they also took on the role of motivator and trainer in preparing the students for the test. The teachers reported student attitudes as a constraint, not practical constraints like class size and time. Apparently, students were very resistant towards English and were not motivated to put in any extra effort for the subject. Finally,

teacher training was offered in the study in order to standardize procedures in conducting and preparing for the test, and hence, to get reliable and valid scores.

Alastrué & Pérez-Llantada (2010) researched teachers' attitudes towards the assessment of oral skills in the new Selectividad English exam in 2004 and 2009 in Spain. The participants were almost 200 Bachillerato teachers from the Community of Aragón. The oral test would be influential at University entrance exam and represent 30% of the total mark. The data retrieved from longitudinal surveys showed that more than 80% of English language teachers in secondary schools felt that they could not devote enough time to the teaching of oral skills. The main reason was attributed to lack of curriculum time. In spite of this, almost 80% of the teachers felt that oral skills should be included and assessed in the University entrance exam. The survey further revealed that secondary school teachers were very much aware of the need for assessing the oral skills of the foreign language as a means of fostering the students' competence in these skills. However, they were dissatisfied with their students' lack of interest.

Höl (2010) tried to explore the attitudes and perceptions of 210 students and 32 instructors towards the speaking test at a School of Foreign Languages in Pamukkale University. The data revealed that most of the students had no experience of any speaking test before, and they were not proficient enough to express themselves even at the elementary level in the target language. Therefore, most of the students considered the speaking test as the most difficult and stressful test. On the other hand, the instructors also thought that the speaking test was the most difficult one to apply and assess, however, the scale and rubrics were sufficient enough to assess the students' oral performance.

Lee (2010) inquired into the current status of classroom speaking assessment and its effectiveness in secondary schools in South Korea. The data were gathered from 51 Korean English teachers for questionnaires and six of them for interviews. The results revealed that classroom speaking assessment broadly employed performance-based tasks, and teachers offered informative feedback to students in the form of criterion descriptions and scores. Additionally, teachers had an overall pessimistic attitude towards the positive effects of such testing on teaching and learning. It was evidence that there was a need for improvements in order to facilitate better learning outcomes in the classroom. Accordingly, the study provided some suggestions for an improvement of current practices such as teacher training, cooperation with an English native teacher, and downsizing the number of students per class.

Duran (2011) investigated teachers' and students' beliefs about teaching and testing speaking and their perceptions of the washback effects of classroom-based speaking tests.

The study was administered with 307 preparatory class intermediate level students and 45 English language instructors at Akdeniz University School of Foreign Languages in Antalya, Turkey. The study revealed both teachers' and students' positive attitudes towards the importance of teaching and testing speaking. The teachers considered that speaking skill could be measured accurately but it was difficult to test. Teachers and students believed that speaking tests improved speaking skills of the students as a washback effect. However, washback could not be seen on what teachers taught, what students learnt, and what they did in classes.

Majid, Samad, Muhamad and Vethamani (2011) sought the implementation of the school-based Oral English test (OET) with 30 students and 14 teachers in selected Maktab Rendah Sains MARA schools in Malaysia. The results of the data revealed that both students and teachers wanted to change the implementation of the OET. While the students desired more help from the teachers, the teachers were concerned with the frequency of the test and tasks. The study highlighted the test which would change the students' learning and test-taking strategies. Besides, the scoring rubrics would model the kinds of instructional tasks and processes that teachers should use in their classrooms.

## **2.12. Conclusion**

This chapter reviewed the literature on theoretical framework of speaking assessment, definitions and importance of language tests, history of testing speaking as a second language, positive and negative aspects of speaking tests, test specifications, techniques for testing speaking, test administration, rater and interlocutor training, the CEFR, rating scales, and related research studies. The next chapter will introduce the methodology of the study. Research design, setting, participants, instruments, data collection procedures and data analysis will be presented in detail.

## **CHAPTER THREE: METHODOLOGY**

This chapter comprises six subsections: research design, setting, participants, instruments, data collection procedures and data analysis. Firstly, the design of the study will be remarked. Next, setting of the study will be clarified. Later, participants of the study will be specified. After that, instruments and data collection procedures will be explained. Finally, information on the data analysis will be provided.

### **3.1. Research Design**

The study was designed to set how English language teachers assess their students' speaking skills in Anatolian high schools. In addition, the study attempted to investigate the 9<sup>th</sup> grade students' and their teachers' attitudes toward speaking tests. Both quantitative and qualitative research methods, namely mixed method were conducted in this study to provide an enhanced understanding. Qualitative and quantitative research corroborate each other and provide different perspectives. According to Creswell & Clark (2011), “qualitative data provide a detailed understanding of a problem while quantitative data provide a more general understanding of a problem” (p. 8). However, each of the research methods has its own limitations, and the limitations of one method can be neutralized by the strengths of the other method.

Creswell & Clark (2011) classify mixed method designs into those: the convergent parallel design, the explanatory sequential design, the exploratory sequential design, and the embedded design. This study was designed as the convergent parallel design prioritizing both quantitative and qualitative methods equally. Qualitative and quantitative data were collected in a concurrent time, analyzed independently and interpreted dependently in the convergent parallel design. To illustrate, I surveyed the 9<sup>th</sup> grade students and their teachers, and I also interviewed both in the same phase of the research. Then, I analyzed the questionnaires quantitatively and the interviews qualitatively. Finally, I compared/contrasted and interpreted results of the qualitative and quantitative data. The purpose of this design was “to obtain different but complementary data on the same topic” (Morse, 1991, p. 122).

### **3.2. Setting**

After pilot study was performed in two Anatolian high schools, the main study was carried out in six Anatolian high schools, in Denizli, in the spring term of 2016-2017 academic year. I tried to include all Anatolian high schools in Denizli in my study. But speaking skills of the students were not assessed in all high schools even though teachers



knew speaking tests were compulsory in high schools. So I could determine just six Anatolian high schools for the main study and two Anatolian high schools for the pilot study. Luckily, the schools were classified in terms of their students' placement scores in Temel Eğitimden Ortaöğretime Geçiş (TEOG) [Transition from Primary to Secondary Education] exam as high level (two of them), middle level (two of them), and low level (two of them). The schools categories are clearly visible in Table 3.1 for main study and in Table 3.2 for pilot study.

Table 3.1

*The Schools Categories in Terms of Students' Placement Scores in Main Study*

<b>Schools Categories</b>	<b>Base score ranges of the students in TEOG</b>
High level	474.7789 - 462.8894
Medium level	457.9642 - 432.6149
Low level	428.4378 - 390.5136

*Note:* It is retrieved from <https://www.personelmebhaber.net/son-dakika/teog-denizli-liseleri-taban-puanlari-2015-2016-h30773.html>.

Table 3.2

*The Schools Categories in Terms of Students' Placement Scores in Pilot Study*

<b>Schools Categories</b>	<b>Base scores of the students in TEOG</b>
Medium level	435.0340
Low level	384.3819

*Note:* It is retrieved from <https://www.personelmebhaber.net/son-dakika/teog-denizli-liseleri-taban-puanlari-2015-2016-h30773.html>.

In Anatolian high schools, the classes are composed randomly. Each class consists of maximum 34 students in these schools. Additionally, the high schools have the same curriculum which focuses on all four skills. The students revise and improve the CEFR A1 level, and after that, they proceed through the CEFR A2 and B2 levels (Ministry of National Education, 2014). In assessment process, achievement tests are written and administered by English language teachers in Anatolian high schools. It is compulsory to do pencil-and-paper tests at least twice in each academic term (Regulation for Secondary Education Institutions, 2016). Besides, a performance test has to be applied for speaking skills in each academic term. Although speaking tests have been compulsory in high schools in Turkey since 2014 (Regulation for Secondary Education Institutions, 2014), there is no standardization for speaking tests. Test results just affect students' achievement in English classes. As a matter of fact, "the primary role of an achievement test is to determine whether course objectives have been met –and appropriate knowledge and skills acquired- by the end of a period of instruction" (H.D. Brown, 2004, p. 48).

### 3.3. Participants

The participants were the 9<sup>th</sup> grade students and their teachers from six Anatolian high schools in Denizli. The 9<sup>th</sup> grade students were selected because they had six hours of English classes per week while the other graders had four hours of English classes (Ministry of National Education - Board of Education, 2014). Totally 380 people from Denizli participated in this research. 358 students and 22 teachers were surveyed. Subsequently, 112 students out of 358 students and 13 teachers out of 22 teachers were interviewed.

#### 3.3.1. Students

358 students from the CEFR A1 & A2 level (basic user) were randomly selected as participants for quantitative research. 112 students out of 358 students participated voluntarily in qualitative research. 222 (62%) students out of 358 students were female, and 136 (38%) students were male (see Figure 3.1).

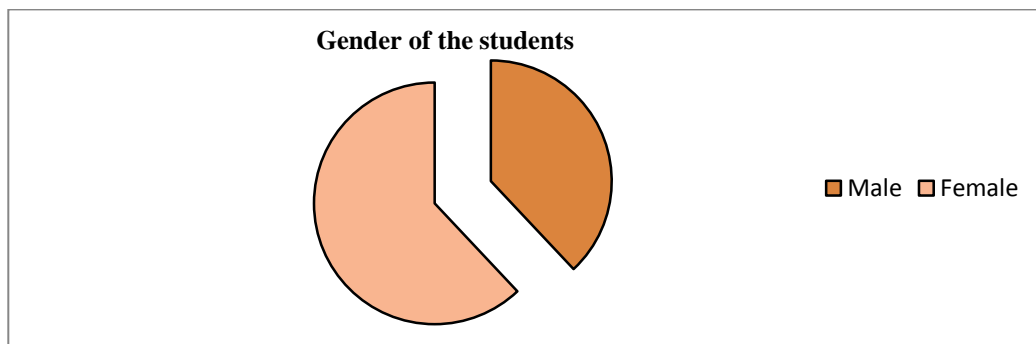


Figure 3.1. The demonstration of gender of the students

The age of the students ranged from 14 to 16 (see Table 3.3). Most of them were 15 years old.

Table 3.3

*Age Range of the Students*

Age	Frequency	Percent
14	42	11.8
15	273	76.2
16	43	12
<b>Total</b>	<b>358</b>	<b>100</b>

#### 3.3.2. Teachers

22 English language teachers were randomly selected as participants for quantitative research. 13 teachers out of 22 teachers participated voluntarily in qualitative research. 17 (77.3%) teachers out of 22 teachers were female, and five (22.7%) teachers were male (See Figure 3.2).

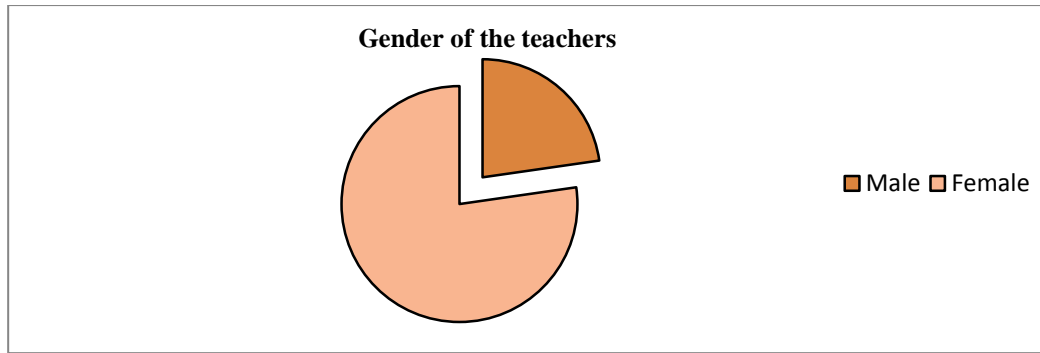


Figure 3.2. The demonstration of gender of the teachers

The age range of the teachers is demonstrated in Table 3.4. The teachers were generally between 31 and 40.

Table 3.4

*Age Range of the Teachers*

Age	Frequency	Percent
21-25	1	4.5
31-35	7	31.8
36-40	7	31.8
41-45	2	9.1
46-50	5	22.7
<b>Total</b>	<b>22</b>	<b>100</b>

While 17 (77.3%) teachers graduated from English Language Teaching, five (22.7%) teachers graduated from English Language and Literature or American Culture and Literature. In addition, only two (9.1%) of the teachers held a master's degree, 19 (86.4%) of them had merely bachelor's degree, and one of the teachers did not reply this item. Experience of the teachers is illustrated in Table 3.5. Their experience was parallel to their age, and it was generally between 11 and 20.

Table 3.5

*Experience of the Teachers*

Experience	Frequency	Percent
1-5	1	4.5
6-10	2	9.1
11-15	9	40.9
16-20	6	27.3
21-25	3	13.6
more than 26	1	4.5
<b>Total</b>	<b>22</b>	<b>100</b>

### 3.3.3. Participants for the Pilot Study

On 19 January 2017, the pilot study was conducted in two Anatolian high schools in Denizli. The aim was to get a fix on shortcomings of the instruments. As in the main study, randomly selected 31 students and 10 teachers took part in the pilot study. 31

students were from the CEFR A1 and A2 levels as in the main study. 19 (61.3%) of the students were female, and 12 (38.7%) were male. Their age range is presented in Table 3.6, and most of the students were 15 years old.

Table 3.6

*Age Range of the Students in Pilot Study*

<b>Age</b>	<b>Frequency</b>	<b>Percent</b>
14	7	22.6
15	20	64.5
16	4	12.9
<b>Total</b>	<b>31</b>	<b>100</b>

Seven (70%) of the teachers were female and three (30%) of them were male. Their age range is shown in Table 3.7, and most of the teachers were between 26 and 40.

Table 3.7

*Age Range of the Teachers in Pilot Study*

<b>Age</b>	<b>Frequency</b>	<b>Percent</b>
26-30	2	20
31-35	2	20
36-40	3	30
51-55	1	10
56-60	1	10
No answer	1	10
<b>Total</b>	<b>10</b>	<b>100</b>

The teachers graduated from different departments as seen in Table 3.8. However, none of the teachers held a master's degree.

Table 3.8

*Diversity in the Teachers' Bachelor of Art in Pilot Study*

<b>Bachelor of art</b>	<b>Frequency</b>	<b>Percent</b>
English Language Teaching	6	60
English Language and Literature / American Culture and Literature	1	10
Translation and Interpretation	1	10
Other	2	20
<b>Total</b>	<b>10</b>	<b>100</b>

In Table 3.9, experience of the teachers is clearly visible. Generally, year range of their professional life was between 6-10 and 16-20.

Table 3.9

*Experience of the Teachers in Pilot Study*

<b>Experience</b>	<b>Frequency</b>	<b>Percent</b>
1-5	1	10
6-10	3	30
11-15	1	10
16-20	3	30
21-25	1	10
more than 26	1	10
<b>Total</b>	<b>10</b>	<b>100</b>

### 3.4. Instruments

Both quantitative and qualitative data collection tools were adopted in the study. In the quantitative part of the study, the questionnaires for both teachers and students were applied (see Appendices C and D). In the qualitative part of the study, the interviews for both were the main source to enrich data (see Appendices E and F). Fulcher (2014) also believes that collecting data through the questionnaires and the interviews is an important method in finding out students' and teachers' preferences and opinions.

The instruments in this study were produced by analyzing some theses like Duran (2011), Güllüoğlu (2004), Höl (2010), Lozovska-Güneş (2010), Önal (2010). They were held in both Turkish (mother tongue) and English. Turkish version of the instruments was offered in Appendices G, H, I, J. They were optional for teachers but students were introduced in Turkish to prevent any misunderstanding or communication breakdown. Four experts in ELT department, two English language teachers and three Turkish teachers were consulted in order to ensure face and content validity of the instruments, and then the instruments took their final version.

In the questionnaires, first, the aim and importance of the study were explained for teachers and students. Then, demographic information was demanded. Demographic information provided background about the students (age, gender) and the teachers (age, gender, educational background, experience). The rest of the questionnaires involved two parts. Part 1 was composed of five-point Likert scale (1: strongly disagree; 2: disagree; 3: partly agree; 4: agree; 5: strongly agree), and Part 2 embodied multiple choice items and open-ended items asking participants' comments.

In teachers' questionnaires, there were 52 items in Part 1. Part 2 embraced 14 multiple choice items, and also the teachers could add their extra comments. The main sections of the teachers' questionnaire are as follows:

- The items related to how speaking skills are assessed: 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 22, 23, 24, 25, 26, 32, 33 and 36 in Part 1; 2, 3, 4, 5, 6, 7, 8, 10, 11 and 12 in Part 2 (See Table 3.10).

Table 3.10

*The Items Related to How Speaking skills Are Assessed in Teachers' Questionnaire*

<b>Part 1</b>
4. I plan the process of the speaking test with my colleagues at the beginning of the term or academic year.
5. I prepare a rubric or a checklist for the speaking test.
6. My colleagues and I agreed on rubric items together.
7. I let the students know what testing procedure will be adopted prior to a speaking test.
8. I declare rubric items and marks for each item before a speaking test.
9. I give extra materials to the students before the speaking test so that they can study better.
10. I give the necessary time for the students in order to cram for the speaking test by announcing the time of it in advance.
12. I tested the students in a room which is quiet.
13. I tested the students in the classroom with the other students.
14. The classroom was noisy during the speaking test.
16. I just listened to the students and gave their marks as an assessor during the speaking test.
17. I both asked questions to the students and gave their marks as both assessor and interlocutor during the speaking test.
18. I combined speaking skill with another skill (listening, reading or writing).
19. The time was enough for each student.
22. I tolerated their mistakes not to demotivate the students during the speaking test.
23. I supported the students with my positive behavior during the speaking test.
24. I behaved in a biased way for some students during the speaking test.
25. I used a checklist or a rubric to assess their performance.
26. I recorded the students' performances.
32. I gave each student's mark during the speaking test.
33. I gave each student's mark at the end of the speaking test after all students had finished it.
36. I was objective in assessing the performance of the students.
<b>Part 2</b>
2) Did you plan timing before a speaking test? How much time did you allocate for each student during a speaking test?
3) Which one or ones of them affected you while preparing a speaking test? You can circle more than one choice.
4) How did you support your students to improve their speaking skills and practice or get prepared for speaking tests? You can circle more than one choice.
5) How many times do the students take speaking test in a term?
6) How many teachers test speaking skill of the students?
7) When did you test the students? You can circle more than one choice.
8) How did you test your students? You can circle more than one choice.
10) What kind of materials did you use during the speaking test? You can circle more than one choice.
11) What type of a rating scale did you prefer while testing students' speaking skills? Why?
12) What did you take into consideration related to the following criteria below while testing speaking? You can circle more than one choice.

- The items related the teachers' perception and attitudes towards the speaking test: 1, 2, 3, 11, 15, 20, 21, 27, 28, 29, 30, 31, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51 and 52 in Part 1; 1, 9, 13, 14 in Part 2 (See Table 3.11).

Table 3.11  
*The Items Related the Teachers' Perception and Attitudes towards the Speaking Test*

<b>Part 1</b>
1. I have prepared a speaking test before.
2. I have administered a speaking test before.
3. I am inexperienced with speaking tests.
11. I have some worries about the administration of a speaking test before.
15. It is better to assign only one interlocutor and one assessor as graders.
20. I feel more anxious and nervous during the speaking test in contrast with pen-and-paper tests.
21. The students' anxiety level is very high during the speaking test.
27. The tasks and activities were not above the students' abilities and levels.
28. The tasks and activities were parallel with those in the course book.
29. The tasks and activities were clear and comprehensible.
30. The tasks and activities were from daily life.
31. Some students whose English level is high cannot perform well due to their test anxiety during the speaking test.
34. Rubric items for the speaking test were clear.
35. The rubric for the speaking test was satisfactory to assess them effectively.
37. I think I am a reliable rater.
38. I know that the speaking test is compulsory at high schools.
39. Speaking skills can be assessed accurately.
40. The speaking test is necessary to see the students' speaking level.
41. The speaking test encourages students to use their speaking ability.
42. The speaking test helps the students to notice the weaknesses in their speaking performances.
43. Trying to test speaking skill is a waste of time.
44. If speaking skills were not tested, I would not spend so much time on teaching the speaking skill.
45. It is difficult for me to prepare the speaking test.
46. It is stressful to administer the speaking test for me.
47. All teachers should use the rubric developed and imposed by the Ministry of National Education.
48. It is difficult to prepare rubric for the speaking test.
49. Rubric is useless in the speaking test.
50. It is difficult to be objective as marking the speaking test for me.
51. I believe that all the teachers at my school are objective.
52. All teachers should be trained through in-service training on how speaking skill can be tested.
<b>Part 2</b>
1) Have you ever taken part in interlocutor training sessions before speaking tests? Where?
9) What kind of tests do you think are more advantageous to use while testing speaking skills? You can circle more than one choice. Why?
13) What are great challenges in the assessment of speaking skills in your opinion? You can circle more than one choice.
14) How should speaking skills be tested?

In students' questionnaires, Part 1 was comprised of 44 items. Part 2 embraced eight multiple choice items, and also the students could add their comments. The main sections of the students' questionnaire are as follows:

- ❖ The items related to how speaking skills were assessed in students' questionnaire: 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 16, 17, 18, 26, 31, 32 and 33 in Part 1; 2, 3, 4, 5 and 6 in Part 2 (See Table 3.12).

Table 3.12

*The Items Related to How Speaking Skills Are Assessed in Students' Questionnaire*

<b>Part 1</b>
1. I was informed about testing procedure prior to the speaking test.
2. I was informed about how my performance would be assessed prior to the speaking test.
3. My teachers gave me extra materials to be able to study better before the speaking test.
4. I had sufficient time to study for the speaking test.
6. I had no idea about the grading of the speaking test.
7. I took speaking test alone in the classroom.
8. I took speaking test with my one or two friends as pair or group in the classroom.
9. As I was taking speaking test, the other students were also in the classroom.
10. It discomforted me in terms of attention, tension, noise, etc. because the other students were also in the classroom as I was taking speaking test.
11. My teachers' attitude toward me influenced my performance negatively during the speaking test.
12. My teachers gave me sufficient time to think about my performance during the speaking test.
13. I had difficulties in understanding pronunciation of my teacher.
16. My teachers encouraged and supported me to perform better during the speaking test.
17. My teachers gave me feedback related to my performance in the speaking test.
18. The speaking test was recorded.
26. The time was not sufficient for each student in speaking test.
31. My teachers were objective at scoring.
32. My teachers assessed my performance on the basis of a rubric or a checklist.
33. The grade I got from speaking test correctly reflected my speaking ability.
<b>Part 2</b>
2) What did you do to get ready for the speaking test? You can circle more than one choice.
3) How many teachers were there in the speaking test?
4) When do you take your speaking test at your school? You can circle more than one choice.
5) Did your teachers use any material during the speaking test?
6) Which activities did you do in this speaking test? You can circle more than one choice.

- ❖ The items trying to find out students' perception and attitudes towards the speaking test: 5, 14, 15, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44 in Part 1; 1, 7 and 8 in Part 2 (See Table 3.13).

Table 3.13

*The Items Related the Students' Perception and Attitudes towards the Speaking Test*

<b>Part 1</b>
5. I was inexperienced with how I should get ready for a speaking test.
14. It is hard to express myself clearly during the speaking test.
15. My personal characteristics such as coyness, timidity, nervousness, etc. stop me from speaking English during the test.
19. The tasks and activities were loud and clear in the speaking test.
20. The speaking test was parallel with class activities.
21. The tasks and activities were from daily life in the speaking test.
22. Some tasks and activities were non-class and unexpected in the speaking test.
23. The tasks and activities were difficult.
24. Speaking activities during lessons were not sufficient.
25. I wanted to do more speaking activities in class after this test.
27. I notice my weaknesses in speaking after speaking test.
28. I feel that I need to study English harder after speaking test.
29. I performed better than I expected in speaking test.
30. I notice my strengths in speaking after speaking test.
34. Speaking test is essential to identify my speaking level.
35. I feel more relaxed in speaking tests in contrast with pen-and-paper tests.
36. If I did not have to take a speaking test, I would not take part in speaking activities in the lessons.
37. Speaking test is useful to improve my English.
38. I can also use many of the things in lessons which I have studied for the speaking test.



39. If I did not have to take a speaking test, I would not spend so much time to improve my speaking skill.
40. I think I need to take speaking tests more frequently.
41. Speaking test makes me nervous.
42. Speaking test is the most difficult test.
43. Speaking test is a test that I find myself the most unsuccessful.
44. It is waste of time to assess speaking skill.

---

**Part 2**

---

- 1) Have you ever taken a speaking test before? Where?
- 7) Which difficulties did you have in speaking test? You can circle more than one choice.
- 8) In your opinion, which criteria should be taken into consideration to assess a student in a speaking test? You can circle more than one choice.
- 

In the interviews, five questions were asked to both the teachers and the students who had already answered the questionnaires. The aim was to triangulate the data from questionnaires and interviews. The questions were related to opinions, manners and preferences of the teachers and students for speaking tests. Their possible suggestions were also asked. Reliability issue of the speaking tests was examined. The interview questions for both the teachers and the students are provided in Table 3.14 and Table 3.15.

Table 3.14

*Teacher Interview Questions*

---

<b>Interview Questions</b>
1. What aspects of the speaking test do you like most?
2. What aspects of the speaking test do you like least?
3. Do you have any suggestions on how to improve the procedure of assessment of speaking?
4. Can you briefly describe a good quality speaking test procedure in your opinion?
5. What kind of measures do you take to ensure a high level of reliability?
a. Do you consider the issues of inter and intra-rater reliability when you test speaking?
b. If you give different marks to the same student how do you handle it?

---

Table 3.15

*Student Interview Questions*

---

<b>Interview Questions</b>
1. What aspects of the speaking test do you like most?
2. What aspects of the speaking test do you like least?
3. Are you happy with your grade for the speaking test?
a. If you are not, do you talk about it with your teacher?
b. How does your teacher deal with it?
4. What would you like to change related to speaking test?
5. What can else be done in speaking test in your opinion? What do you suggest for it?

---

### **3.5. Data Collection Procedures**

On 28 December 2016, I requested permission from Pamukkale University to administer data collection procedures. Next, on 12 January 2017, I got permission for data collection from Denizli Directorate of National Education (see Appendix K). Right after, on 19 January 2017, pilot study was carried out in two Anatolian high schools in Denizli. Teachers in the study had assessed students' speaking skill in 2016 autumn term. Depending on this assessment, the students and the teachers shared their experiences and

opinions via questionnaires and interviews in this research. As a result of the data analysis in the pilot study, the instruments were revised and necessary modifications were arranged to improve them.

On 3 March 2017, the main study was conducted in two Anatolian high schools. Next week (10 March 2017), data were gathered in another two Anatolian high schools. Lastly (17 March 2017), the procedure was applied in the other two Anatolian high schools. In brief, I investigated two Anatolian high schools each week. Data collection procedures lasted three weeks in total.

I administered the questionnaires in a one-hour-class session. I informed the participants about the content, objectives, and procedures of the study. Additionally, I assured that data would be kept confidential, and they would not be used for any other purposes. The questionnaires were obtained from the participants. Afterwards, volunteer participants were interviewed individually. The interviews were collected through audio recordings. I adapted the questions and the order of questions according to the answers of the participants during the interviews. I gave examples and explanations for the questions to prevent any misunderstandings and clarify meanings of the questions. I encouraged the participants to give more explanations or examples. The interviews were about five minutes for each participant. At the end of the data collection procedure, both quantitative and qualitative data were compiled.

### **3.6. Data Analysis**

The quantitative data were collected through questionnaires and analyzed by Statistical Package for Social Sciences (SPSS) 22.0. First, the data were calculated whether they were parametric or non-parametric by using one-sample Kolmogorov-Smirnov test, and the result indicated that the data were parametric ( $p < 0.001$ ). Next, independent sample t-tests were used in order to find out whether there was a significant difference between male and female students in regard to their perception and attitudes towards the speaking test. After that, one-way Analysis of Variance (ANOVA) was computed to compare the students' perception and attitudes towards the speaking test among Anatolian high schools in terms of the students' placement scores. Finally, items of the questionnaires were analyzed by using descriptive statistics (frequency, percentage and mean scores).

The qualitative data were collected through interviews. They were recorded and transcribed to analyze content thematically. Content analysis is defined as: "a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon,

2005, p. 1278). First, I defined the coding unit in the process of analyzing the qualitative data. Then, I determined themes and sub-themes in the data. After that, I associated the themes and sub-themes to research questions. According to Zhang & Wildemuth (2005), human coders are prone to fatigue and are likely to make more mistakes during coding. What is more, the coders' understanding of the categories and coding rules may vary subtly over the time (Miles & Huberman, 1994; Weber, 1990). For all these reasons, lastly, the consistency of the coding was rechecked by two English language teachers who were also doing their master of art study.

## CHAPTER IV: RESULTS

This chapter presents interpretation of the results on the basis of all the research questions. Firstly, it will be delineated the way how speaking skills are assessed in Anatolian high schools. Then, students' perception and attitudes regarding the speaking test will be described. Additionally, it will be examined whether gender and schools differences affect students' perception related to speaking tests or not. Finally, teachers' perception and attitudes towards the speaking test will be reflected.

### 4.1. The Way How Speaking Skills are Assessed in Anatolian High Schools

In order to make testing process more explicit, those titles are dealt with: planning for the speaking test, test instruction for students, time allocation, test tasks, materials, testing environment, teachers' roles, teachers' behavior, objectivity in assessment, rating scales and feedback. The items investigating how speaking skills are assessed in Anatolian high schools and their quantitative analysis are presented in the following tables.

#### 4.1.1. Planning for the Speaking Test

Table 4.1

*Descriptive Statistics for Planning Process from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
4. I plan the process of the speaking test with my colleagues at the beginning of the term or academic year.	22	4.36	.84	0	0	18.2	31.8	45.5

The mean score for the item 4 related to planning the process of the speaking test is 4.36 as seen in Table 4.1. Teachers (77.3%) remarked that they planned the process of the speaking test with their colleagues at the beginning of the term or academic year. What is more, teachers (81.8%) accepted that they planned timing before speaking test (item 2 in Figure 4.1).

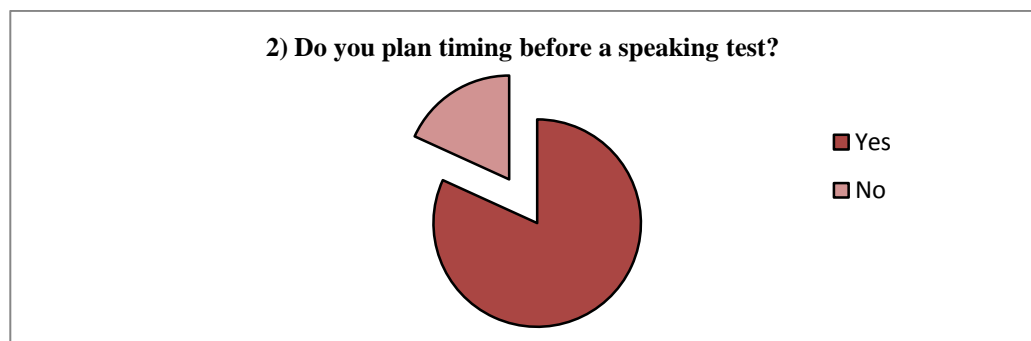


Figure 4.1. The percentage of the teachers who plan timing before a speaking test

Yet, there is no standard for speaking tests in Anatolian high schools. While speaking skill is being assessed in a high school, it may be ignored in another high school. Even some teachers can test differently in the same high school. According to Regulation for Secondary Education Institutions (2014), the teachers just have to do speaking test at least once in each academic term. According to the data, 50% of the teachers did speaking test once in a term while 45.5% of them did it twice in a term (item 5 in Figure 4.2).

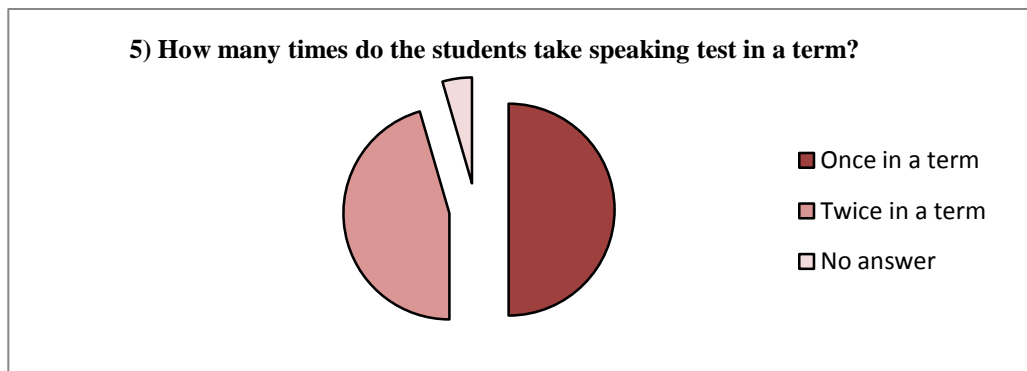


Figure 4.2. The demonstration of the speaking tests' frequency in a term

English language curriculum just recommends that the CEFR should be used as the basis for the curriculum (Ministry of National Education, 2014). Eventually, teachers can plan the procedures of the speaking test individually or mutually with their colleagues. However, team work is a necessity in that standardization occurs among classes in a school. T1 also clarified his/her idea in the interview as follows:

*"...On the behalf of better speaking test, there is a need for standardization..." (Audio Recording: T-1)*

Teachers need to plan speaking test in a detailed way under no circumstances: design test tasks; arrange materials, setting and rubric previously; plan timing so as to overcome possible problems.

#### **4.1.2. Instruction**

Teachers (95.5%) asserted that they declared testing procedure prior to a speaking test (item 7 in Table 4.2). Students also corroborated this item at the percentage of 72.9 (item 1 in Table 4.3). Even so, there were few students who expressed the following statement:

*"We were sometimes assessed without being informed. In that case, I could not speak clearly due to my nervousness" (Audio Recording: S-34/midlevel).*

Table 4.2

*Descriptive Statistics for Instruction from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
7. I let the students know what testing procedure will be adopted prior to a speaking test.	22	4.77	.52	0	0	0	27.3	68.2
8. I declare rubric items and marks for each item before a speaking test.	22	4.59	1.09	4.5	4.5	0	13.6	72.7

Furthermore, teachers (86.3%) indicated that they explained rubric items and marks for each item before a speaking test (item 8 in Table 4.2). Whereas 24.1% of the students (item 6 in Table 4.3) reflected that they had no idea about the grading of speaking test, 69% of the students (item 2 in Table 4.3) approved the teachers, and they pointed out that they were informed about how their performance would be assessed prior to the speaking test. Nonetheless, some students notified their experiences in the interviews like that:

*“I am not happy with my grade for the speaking test because my teacher has not informed us about marks for criteria such as grading for pronunciation, fluency or mimics. If I had known it, I could have studied harder” (Audio Recording: S-74/midlevel).*

*“My teacher scores after his/her own heart. S/he does not share the rating scale with us. We cannot discuss our marks after the speaking test. I do not know why I failed...” (Audio Recording: S-58/low level).*

Table 4.3

*Descriptive Statistics for Instruction from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
1. I was informed about testing procedure prior to the speaking test.	358	3.93	1.16	6.4	5.9	14.5	34.9	38
2. I was informed about how my performance would be assessed prior to the speaking test.	358	3.85	1.21	7.5	6.4	17	31	38
6. I had no idea about the grading of speaking test.	358	2.45	1.37	32.7	26.3	16.2	14	10.1

It is crucial to inform students about testing and grading procedure of speaking tests because they feel more secure and confident when students know how they will prepare for the speaking test, what they will do during test, what criteria will be applied, and how they will be assessed. In addition, they plan their studying depending upon this informing. The data revealed that studying methods of the students for the speaking test (item 2 in Part 2 from students' questionnaire) were as follows:

- They followed the lesson carefully.
- They crammed for the test on their own or with their friends.
- They got professional help (language course or English teacher).

- They practiced conversation in their daily life.
- They listened to music, followed series, and played computer games in English.
- They went abroad.
- They chatted on social media in English.

There was no accumulation on any alternative for studying methods of the students. Notwithstanding, mostly preferred alternative was to cram for the test individually.

#### 4.1.3. Time Allocation

In line with Regulation for Secondary Education Institutions (2014), teachers have to determine and announce the date of the speaking test at the beginning of each academic term.

Table 4.4

##### *Descriptive Statistics for Time Allocation from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
10. I give the necessary time for the students in order to cram for the speaking test by announcing the time of it in advance.	22	4.59	.73	0	0	13.6	13.6	72.7
19. The time was enough for each student.	22	4.36	.90	0	4.5	13.6	22.7	59.1

Teachers (86.3%) believed that they gave necessary time for the students in order to cram for the speaking test by announcing the date and time of it in advance (item 10 in Table 4.4). Students (65.6%) also thought that they had sufficient time to study for the speaking test (item 4 in Table 4.7). Even so, 14.5% of the students contradicted this idea, and verbalized it in the interviews as follows:

*“We did not have enough time to study. As a result, I could not speak as requested, and my sentences were not accurate as I wished. If I had had more time, I would have studied harder” (Audio Recording: S-101/midlevel).*

*“I am not happy with my grade for the speaking test because my teacher announced the date of the speaking test just a day before. We did not have necessary time to study, as a result, this condition affected my mark” (Audio Recording: S-77/midlevel).*

*“The negative aspect of the test was that we had less time to get ready” (Audio Recording: S-73/high level).*

The items 7 and 4 of the questionnaires tried to reveal when speaking tests were done. The results indicated that 50% of the teachers (item 7 in Table 4.5) tested the

students during English lessons, and 72.3% of the students also affirmed it (item 4 in Table 4.6).

Table 4.5

*Test Time from Teachers' Perspective*

<b>7) When do you test the students? You can circle more than one choice.</b>		
<b>Option Responses</b>	<b>Frequency</b>	<b>Percent</b>
a) During the class time	2	9.1
b) In my free time	2	9.1
d) During English lessons	11	50
No answer	1	4.5
<b>Total</b>	<b>16</b>	<b>72.7</b>

*Note:* 6 teachers out of 22 circled more than one choice.

Table 4.6

*Test Time from Students' Perspective*

<b>4) When do you take your speaking test at your school? You can circle more than one choice.</b>		
<b>Option Responses</b>	<b>Frequency</b>	<b>Percent</b>
a) During the class time	31	8.7
b) In my free time	1	.3
c) After school time	1	.3
d) During English lessons	259	72.3
e) In my teacher's free time	25	7
f) Other	7	2
No answer	5	1.4
<b>Total</b>	<b>329</b>	<b>92</b>

*Note:* 29 students out of 358 circled more than one choice.

What is more, regarding the time, teachers generally allocated minimum two minutes and maximum 15 minutes for each student during the speaking test (item 2 in Part 2 from teachers' questionnaire). They (81.8%) accepted that they gave enough time for each student (item 19 in Table 4.4). Students also assented it at the percentage of 75.2 (item 26 in Table 4.7). Moreover, students (59%) stated that they had sufficient time to think about their performance during the speaking test (item 12 in Table 4.7). However, some students reported their complaint about time allocation in the interviews like that:

*"While my teacher was putting pressure on me about time, I failed because I could not overcome my stress" (Audio Recording: S-97/low level).*

*"Time could be longer..." (Audio Recording: S-14/high level)*

*"...My teacher tries to test 30 students in two class hours (80 minutes). Accordingly, each student has two or three minutes, and they cannot speak as they wish. We cannot speak fluently anyway, and it is getting worse when we try to be fast" (Audio Recording: S-58/low level).*



Table 4.7

*Descriptive Statistics for Time Allocation from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
4. I had sufficient time to study for speaking test.	358	3.81	1.18	6.1	8.4	19.8	29.6	36
12. Before my performance, my teachers gave me sufficient time to think about it in the speaking test.	358	3.63	1.24	8.9	8.4	23.7	29.1	29.9
26. The time was not sufficient for each student in speaking test.	358	2.10	1.25	39.4	35.8	10.6	5	8.4

If teachers do not plan time allocation well, students cannot find necessary time for studying, and it affects their performance, motivation, mark, etc. Teachers should be attentive especially for crowded classes, too. The more students take speaking test, the longer it will last. Besides, it will be tiring for teachers, and probably it will influence rater reliability.

**4.1.4. Test Tasks**

According to the data, teachers designed a speaking test under the influence of the curriculum, English course book, students' level and interest, daily social issues related to students' ages and current interests, and affairs in students' real life (item 3 in Part 2 from teachers' questionnaire).

Students asserted that they did these activities: making a presentation on a topic, talking about pictures, introducing yourself and your family, answering questioning randomly from question pouch, answering questions given before the test, acting out dialogues with another classmate, discussing in groups, role-playing, talking about video / film (character, pilot, setting, etc.), talking about something he listens or reads (item 6 in Table 4.8).

Table 4.8

*Descriptive Statistics for Test Activities from Students' Perspective*

<b>6) Which activities did you do in this speaking test? You can circle more than one choice.</b>			
	<b>Option Responses</b>	<b>Frequency</b>	<b>Percent</b>
a)	Making a presentation on a topic	98	27.4
b)	Talking about pictures	6	1.7
c)	Introducing yourself and your family	10	2.8
d)	Answering questioning randomly from question pouch	4	1.1
e)	Answering questions given before the test	4	1.1
f)	Acting out dialogues with another classmate	5	1.4
g)	Discussing in groups	5	1.4
h)	Role-playing	3	.8
i)	Talking about video / film (character, pilot, setting, etc.)	4	1.1
j)	Talking about something he listens or reads	21	5.9
k)	Other	17	4.7
	No answer	6	1.7
<b>Total</b>		<b>183</b>	<b>51.1</b>

*Note:* 175 students out of 358 circled more than one choice.

Although there was no agreeable answer, presentation (27.4%) was the most favorite task as seen in Table 4.8. Topic choice is a crucial issue in presentation tasks. Following statements from the interviews indicated that the teachers cared the students' levels and concerns:

*I think it is important to select appropriate topic in test tasks. Primarily, topics must catch students' attention. They must encourage and motivate students to speak. The topics ought to be pertinent to students' age and interest in order to be personalized (Audio Recording: T-1).*

However, it is not possible to make every student glad. So some students mentioned their discontent as follows:

*"In my opinion, topic of the tasks should have been from daily life. My topic was relevant to 'friendship', and it was okay. But my friend's topic was relevant to 'forgetfulness', and it was really difficult to talk..." (Audio Recording: S-4/high level)*

*"Topics were so classic that I was bored" (Audio Recording: S-28/low level).*

*"I suggest that topics should be more varied because the topic was only students' daily routine" (Audio Recording: S-95/high level).*

*"It was challenging to talk about a topic which was settled by my teacher, and it was stressful not to be able to select topic on my own. If I had talked about familiar topic as I wished, I would have felt more relaxed" (Audio Recording: S-7/midlevel).*

Students can select topics themselves for an interactive test. Additionally, students' responses in the interviews delineate how teachers had better select topics of the test tasks. Topics should be meaningful for students because they influence students' performance and attitudes towards speaking tests. For example, if students are familiar with topics, their speech rate and span will increase. Furthermore, topics should be appropriate to students' age, gender, interest, background, culture, etc. In the contrary case, students cannot reflect their speaking abilities correctly. Ultimately, topics make a significant contribution to students' speech in speaking tests.

**4.1.4.1. Individual or paired/group test tasks?** There is a divergence between the students' and the teachers' responses in our data. Teachers sustained that they (40.9%) tested students individually, in pairs and groups (item 8 in Table 4.9).

Table 4.9

*Teachers' Preferences for Test Tasks*

<b>8) How do you test your students? You can circle more than one choice.</b>		
<b>Option Responses</b>	<b>Frequency</b>	<b>Percent %</b>
a) Individually	9	40.9
a) Individually & b) In pairs	2	9.1
a) Individually & c) In groups	1	4.5
d) All of them	9	40.9
No answer	1	4.5
<b>Total</b>	<b>22</b>	<b>100</b>

However, students (78.2%) objected that they took the speaking test with their one or two classmates as a pair or group in the classroom (item 8 in Table 4.10). Eventually, most of the students were assessed individually. Yet, some students wanted to be tested in pairs or groups:

*"I would like to be tested in groups for reduction of my stress" (Audio Recording: S-55/high level).*

*"I would like to be tested in groups of five or six students" (Audio Recording: S-78/high level).*

Table 4.10

*Descriptive Statistics for Test Tasks from Students' Perspective*

<b>The items</b>	<b>N</b>	<b>Mean</b>	<b>Sd</b>	<b>S.D.</b>	<b>D.</b>	<b>P.A.</b>	<b>A.</b>	<b>S.A.</b>
				<b>%</b>	<b>%</b>	<b>%</b>	<b>%</b>	<b>%</b>
8. I took speaking test with my one or two friends as a pair or group in the classroom.	358	1.94	1.44	60.3	17.9	2.8	7.5	10.3

In essence, individual or paired/group testing was not a choice for the students in our study because the teachers did not feel confident about testing in pairs or groups. They thought that pair and group testing could create administration and scoring problems, and they could not handle with these problems alone. At the same time, these kinds of tasks can also help teachers to save time, especially for crowded classes (Egyud and Glover, 2001; Ducasse and Brown, 2009). Notwithstanding, the teachers in our study kept away from pair and group tasks as deduced from the interviews:

*"Maybe I assess students in groups. Yet, I need to give marks to students individually so I am assessing them individually (it is difficult to score each student's performance in groups)..." (Audio Recording: T-2)*

*"In my opinion, discussion is a better way of testing students. But students' achievement levels are not enough for it" (Audio Recording: T-5).*

**4.1.4.2. Discrete or integrated test tasks?** The purpose of the item 18 (in Table 4.11) in the teachers' questionnaire was to determine whether test tasks were discrete or integrated. The mean score for item 18 in Table 4.11 was 3.55. While 40.9% of the teachers indicated that they combined speaking skill with another skill (listening, reading or writing), 27.3% of them disagreed with the item (item 18 in Table 4.11). 27.3% of the teachers partly agreed with the item. Based on the percentages, it could be concluded that teachers were closer to integrated test tasks. In addition, some students suggested it in the interviews:

*“We can talk about any song or film in speaking tests” (Audio Recording: S-25/high level).*

*“We can debate on any book or series during speaking tests” (Audio Recording: S-20/high level).*

Table 4.11

*Descriptive Statistics for Test Tasks from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
18. I combined speaking skill with another skill (listening, reading or writing).	22	3.55	1.29	0	27.3	27.3	13.6	27.3

It is common to combine speaking skill with especially reading and listening skills. Students can read something, then talk about it or present it. However, it is really hard to separate speaking skill from listening skill because they are combined naturally. Teachers must be so careful that listening or reading skills do not interfere with speaking skill while assessing them.

#### **4.1.5. Materials**

There was an inconsistency between the teacher (item 9 in Table 4.12) and the students' responses (item 3 in Table 4.13) in terms of materials provided by the teachers before the speaking test. While 63.6% of the teachers claimed that they gave extra materials to the students before the speaking test, 28.3% of the students confirmed that they took extra materials from their teachers to be able to study better before the speaking test. The difference of the percentages was so high. In the interviews, some students still maintained that they took extra materials, but some not:

*“My teacher informed us topic of the task previously. In line with the topic, we studied it via the internet...” (Audio Recording: S-86/high level)*

“...It would have been better if teachers had given us topics of the tasks to study in advance” (Audio Recording: S-4/high level).

Table 4.12

*Descriptive Statistics for Materials from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
9. I give extra materials to the students before the speaking test so that they can study better.	22	3.95	1.04	0	9.1	27.3	22.7	40.9

Table 4.13

*Descriptive Statistics for Materials from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
3. My teachers gave me extra materials to be able to study better before the speaking test.	358	2.78	1.29	18.7	26.8	24.9	18.2	10.1

In the interviews, I realized that the teachers wanted to assess speaking skill of the students spontaneously but they did not find level of the students adequate for this kind of test. So teachers preferred giving extra materials to students or applying test tasks in class in advance as a solution because students lacked experience, and they were not so aware of the process of speaking test.

During the speaking test, students (42.5%) admitted that their teachers used some materials like interactive board, pencil, paper, pictures, rubric, video, course book, computer, mobile phone for recording (item 5 in Figure 4.3).

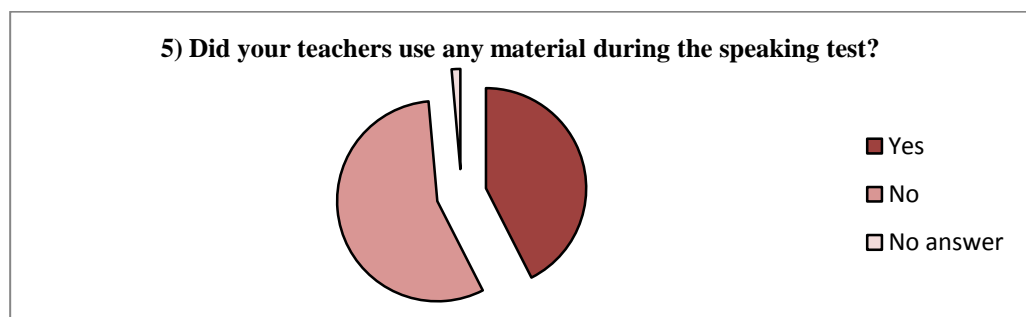


Figure 4.3. The demonstration of the materials usage during the speaking test

Moreover, teachers (item 10 in Part 2 from teachers' questionnaire) remarked that they used such materials: course book, pictures, objects, realia, films, video, EBA software, question pool, topics for discussion, news, magazines, brochures, dialogue for role play, incomplete stories, reading passages, dialogues, topics for presentation.

#### 4.1.6. Testing Environment

Testing environment was a classroom in our study. While 13.1% of the students specified that they took speaking test alone in the classroom (item 7 in Table 4.15), 78.5% of the students noted the contrary (item 9 in Table 4.15). As an implication, test taker's classmates were also in the classroom during the speaking test. Teachers (81.8%) admitted that they tested the students in that environment (item 13 in Table 4.14) because they avoided testing students alone. Teachers believed that if all students witnessed the speaking test, students would appraise the test as reliable and scoring as transparent. Some teachers' views were illustrated as follows:

*“Students’ classmates should witness their performance. If some students share the same topic with their classmates, they can compare one another’s performances” (Audio Recording: T-9).*

*“All students can listen to one another’s performances in classroom...” (Audio Recording: T-4).*

Even so, some teachers were not satisfied with this testing condition:

*“Environment should be convenient. Each student should be tested alone in a quiet class. Teachers should not interfere with students and correct their mistakes during the test” (Audio Recording: T-9).*

*“It is necessary to test students one-to-one. We have to allocate enough time for each student; however, we cannot apply it due to the crowded classes” (Audio Recording: T-7).*

Table 4.14

*Descriptive Statistics for Testing Environment from Teachers’ Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
13. I tested the students in the classroom with the other students.	22	4.23	1.11	4.5	4.5	9.1	27.3	54.5
14. The classroom was noisy during the speaking test.	22	2.82	1.53	22.7	22.7	31.8	0	18.2
12. I tested the students in a room which is quiet.	22	2.82	1.22	18.2	18.2	36.4	18.2	9.1
26. I recorded the students’ performances.	22	2.36	1.00	18.2	40.9	31.8	4.5	4.5

As a solution, teachers could record test process when they tested students alone. However, 9% of the teachers implied that they recorded the students' performances (item 26 in Table 4.14), and 15.7% of the students also reflected so (item 18 in Table 4.15). It was deduced from the data that teachers maintained this administration even though they

were displeased with the existing condition. Moreover, some students were not glad to be tested when other students were in the classroom, either:

*“...Human being can be shy but they (teachers) do not care...” (Audio Recording: S-3/low level).*

*“My classmates laughed at me since I could not speak in public. I disgraced myself” (Audio Recording: S-110/low level).*

*I felt stressful since I was tested in the classroom, and everybody fixed their gaze on me. I wish my teacher had tested me alone. I felt nervous. I forgot some vocabularies, and I had trouble while I was forming a sentence (Audio Recording: S-4/high level).*

On the other hand, some students felt grateful to have a chance to be tested in classroom. They were persuaded that speaking in public would prepare them for a possible job in the future. The students trusted that their self-confidence would increase, and they would perform better as long as they practiced English speaking. Some expressions from these students were like that:

*“Normally, I feel shy when I speak English in classroom but I try to overcome my shyness, and I am getting used to being tested in classroom” (Audio Recording: S-99/midlevel).*

*“I could not reflect myself on the community. It might have been better if I had been tested alone” (Audio Recording: S-1/high level).*

*“It (speaking in public) can benefit us abroad” (Audio Recording: S-76/midlevel).*

*“Speaking tests improve my speaking skill. For example, now, I can talk with a tourist better” (Audio Recording: S-68/high level).*

*“I think we should speak in public more often” (Audio Recording: S-82/high level).*

Arslan (2013) also advocates this idea in his study which investigates pre-service English language teachers' competence in basic elements of spoken communication, and he remarks that students will build their confidence in speaking and lessen their anxieties as long as they practice English speaking.

Table 4.15

*Descriptive Statistics for Testing Environment from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
9. As I was taking speaking test, the other students were also in the classroom.	358	4.19	1.34	10.6	4.2	5.6	16.2	62.3
10. It discomforted me in terms of attention, tension, noise, etc. because the other students were also in the classroom as I was taking speaking test.	358	2.67	1.48	30.2	22.3	16.8	11.5	19.3
18. The speaking test was recorded.	358	1.80	1.37	66.5	14.2	3.4	4.5	11.2
7. I took speaking test alone in the classroom.	358	1.77	1.43	68.7	13.7	2.8	4.2	8.9

18.2% of the teachers (item 14 in Table 4.14) represented that the classroom was noisy during the speaking test while 27.3% of the teachers (item 12 in Table 4.14) advocated that the classroom was quiet during the speaking test. Although the items 14 and 12 were counter-view, both of the percentages (18.2% and 27.3%) were low. Consequently, according to the data, we could not delineate that the classroom was either noisy or quiet. However, 30.8% of the students sustained that they were discomforted in terms of attention, tension, noise, etc. (item 10 in Table 4.15), and they also signified their discomfort in the interviews as follows:

*"I lost my attention, and I could not concentrate on account of noise in classroom" (Audio Recording: S-75/midlevel).*

*"It would have been better if my classmates had been out of the class during speaking test. I got confused by virtue of their noise" (Audio Recording: S-74/high level).*

*"It (speaking test) could be in a quieter and more comfortable place" (Audio Recording: S-65/high level).*

#### **4.1.7. Teachers' Roles: Assessor or Interlocutor?**

The items 16 and 17 (in Table 4.16) of the teachers' questionnaire were expected to ascertain teachers' role as an interlocutor or assessor during the speaking test. Teachers (81.8%) indicated that they adopted both of the teachers' roles: interlocutor and assessor individually (item 17 in Table 4.16). In other words, as students (98%) clarified, there was only one teacher in the speaking test (item 3 in Part 2 from students' questionnaire), and the teachers approved it at the percentage of 90.9%, too (item 6 in Part 2 from teachers' questionnaire).



Table 4.16

*Descriptive Statistics for Teachers' Roles from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
17. I both asked questions to the students and gave their marks as both assessor and interlocutor during the speaking test.	22	4.23	.97	0	9.1	9.1	31.8	50
16. I just listened to the students and gave their marks as an assessor during the speaking test.	22	4.00	1.15	4.5	4.5	22.7	22.7	45.5

In my interviews, I apprehended that teachers did not prefer cooperation in the assessment process; instead, they would rather cooperate in planning and designing test task, materials and rubrics in terms of standardization. But they felt that they were questioned and even judged by their colleagues in the assessment process. As such, cooperation meant intervention in their scoring way for them. Ultimately, teachers would prefer to assess students' speaking skills individually. Yet, it could be distracting and tiring for teachers to both assess and interact. One of the teachers articulated this need like that: *"In the process of evaluation, it can be jury consisting of at least two teachers instead of one teacher..."* (Audio Recording: T-12).

Besides, two of the students explained why they needed both interlocutor and assessor in the interviews:

*"...There must be more than one teacher during assessment. A teacher may assess in a biased way for some students so I would feel more comfortable if I was assessed by more than one teacher"* (Audio Recording: S-7/midlevel).

*"...It can be troublesome for teachers to both interact with and score students during assessment so there may be one or two more teachers to help them"* (Audio Recording: S-84/high level).

According to the data, 68.2% of the teachers verbalized that they adopted merely assessor's role (item 16 in Table 4.16). But there was not another teacher as an interlocutor in the speaking test. The reason for this was the teachers' task preferences. When they selected presentation task as an assessment tool, they considered that there was no need for interaction. Eventually, test takers just made a presentation in the speaking test. Their classmates were passive, and they solely listened to the presentation while the teachers were assessing the test takers. Just so, the mean of item 13 in Table 4.17 was so low: 2.34. Simply 15.4% of the students pointed out that they had difficulties in understanding pronunciation of their teacher because they did not interact. Students searched presentation topic on the Internet, and then they studied on vocabulary knowledge and pronunciation of

unknown vocabularies to memorize them. Lastly, they made their presentation in the speaking test.

Table 4.17

*Descriptive Statistics for Teachers' Roles from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
13. I had difficulties in understanding pronunciation of my teacher.	358	2.34	1.36	31.3	30.4	22.3	8.1	7.3

**4.1.8. Teachers' Behaviors**

Teachers supported their students to improve their speaking skills and get prepared for speaking tests by encouraging and motivating them to speak more often, doing more class activities, suggesting extra sources such as video, film, etc. (item 4 in Part 2 from teachers' questionnaire). Moreover, teachers supported their students with their positive behavior during the speaking test (item 23 in Table 4.18).

Table 4.18

*Descriptive Statistics for Teachers' Behavior from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
23. I supported the students with my positive behavior during the speaking test.	22	4.95	.37	0	0	0	9.1	86.4
22. I tolerated their mistakes not to demotivate the students during the speaking test.	22	4.41	.73	0	4.5	0	45.5	50

In addition, teachers (95.5%) specified that they tolerated the students' mistakes not to demotivate them during the speaking test (item 22 in Table 4.18). Even, a teacher noted his/her support in the interview like that:

*"...I always try to motivate my students; hence, they are getting eager for speaking. Accordingly, they are not afraid of making mistakes, and they do not refrain from their teachers' scoring way" (Audio Recording: T-4).*

However, students had different opinions related to teachers' support to them. Whereas 45% of the students represented that the teachers encouraged and supported them to perform better during the speaking test (item 16 in Table 4.19), 30.2% of them denied their teachers' support and expressed converse in the interview, too:

*"I saw my teacher's derisory glance during the speaking test. She can even scold me when I make a mistake" (Audio Recording: S-9/high level).*

On the other hand, 83.8% of the students opposed that their teachers' attitude influenced their performance negatively during the speaking test (item 11 in Table 4.19). A student justified this idea with following statement:

*"In my opinion, my teacher gives me good marks although she knows that I am unsuccessful"* (Audio Recording: S-36/low level).

Table 4.19

*Descriptive Statistics for Teachers' Behavior from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
16. My teachers encouraged and supported me to perform better during the speaking test.	358	3.20	1.32	13.4	16.8	24.6	28.2	16.8
11. My teachers' attitude toward me influenced my performance negatively during the speaking test.	358	1.78	1.05	51.1	32.7	7.5	4.2	4.5

In my interviews, teachers stated that they generally gave high marks so as to show their support or fix a problem because they viewed marks as a motivation tool. Some examples of the teachers' statements in the interviews:

*"...I do not give very low marks to students; generally I give good marks to motivate them"* (Audio Recording: T-10).

*"I use rubric. But I still give extra points to all students when I have trouble with scoring"* (Audio Recording: T-11).

*"We do not give low marks to anybody in speaking tests. We generally give higher marks than the other test (pencil-and-paper tests)"* (Audio Recording: T-6).

*"I ignore minor mistakes or I do not take very high points off if a student can express himself or herself sufficiently"* (Audio Recording: T-2).

#### **4.1.9. Objectivity in Assessment**

Teachers (95.5%) stated they were objective in assessing performances of the students (item 36 in Table 4.20). What is more, teachers (77.2%) refuted that they behaved in a biased way for some students during the speaking test (item 24 in Table 4.20), and they declared it in the interviews:

*"I do not treat students differently. I design a rubric. All students have the same conditions, and they are subject to this rubric"* (Audio Recording: T-4).

*“It was difficult to be objective although I determined criteria like pronunciation, fluency, etc. in advance, and I used rubric as assessing my students” (Audio Recording: T-8).*

Table 4.20

*Descriptive Statistics for Objectivity from Teachers’ Perspective*

<b>The items</b>	<b>N</b>	<b>Mean</b>	<b>Sd</b>	<b>S.D.</b> <b>%</b>	<b>D.</b> <b>%</b>	<b>P.A.</b> <b>%</b>	<b>A.</b> <b>%</b>	<b>S.A.</b> <b>%</b>
36. I was objective in assessing the performance of the students.	22	4.64	.58	0	0	4.5	27.3	68.2
24. I behaved in a biased way for some students during the speaking test.	22	1.91	1.34	54.5	22.7	9.1	9.1	0

The students (71.5%) also acknowledged that their teachers were objective in scoring (item 31 in Table 4.21). Even so, some students disagreed and verbalized contradictory ideas in the interviews with following statements:

*“...I think my teacher is favoring some students a bit. I would like my teacher to be more objective” (Audio Recording: S-74/high level).*

*“My teacher scores my speaking skill under the influence of my performance during class hours...” (Audio Recording: S-24/high level)*

Table 4.21

*Descriptive Statistics for Objectivity from Students’ Perspective*

<b>The items</b>	<b>N</b>	<b>Mean</b>	<b>Sd</b>	<b>S.D.</b> <b>%</b>	<b>D.</b> <b>%</b>	<b>P.A.</b> <b>%</b>	<b>A.</b> <b>%</b>	<b>S.A.</b> <b>%</b>
31. My teachers were objective at scoring.	358	3.95	1.18	6.4	5.9	14.8	33.2	38.3

Furthermore, some teachers used supportive expressions in the interviews like that:

*“...If a student generally takes high marks but s/he has problem with speaking tests, I tolerate the student in speaking tests, and I give easier tasks to the student on behalf of his/her motivation” (Audio Recording: T-6).*

*“I try to score for the benefit of the students. Namely, even though a student does not speak in spite of my encouragement, I score the student by considering his/her other marks. Even my students generally say that I give them high marks though they do not talk (in the speaking test)” (Audio Recording: T-5).*

*“I determined criteria of the rubric, and added students’ performance during class as a criterion” (Audio Recording: T-12).*

The responses given about the objectivity of the tests by the teachers in the questionnaire and the interviews were not consistent. In my interviews, most teachers had

trouble with objectivity, and quantitative data did not reflect reality. Teachers could not discriminate between students' test performance and class performance. They labeled their students on their mind according to their performance during class hours and especially their previous marks in their pencil-and-paper tests. For example, if a student got 50 from pencil-and-paper test, teachers would give nearly 60 or 40 depending on the student's performance. Ultimately, the teachers assessed their students via their bias. Although they relied on their objectivity, teachers went on wearing blinkers.

#### 4.1.10. Rating Scales

Though rubric is compulsory, there is no standard rubric for speaking tests in high schools. Teachers can benefit from CEFR, design themselves, or get a ready-made rubric. 86.3% of the teachers pointed out that they prepared a rubric or a checklist for the speaking test (item 5 in Table 4.22). Moreover, 63.6% of the teachers reported that they agreed on the rubric items with their colleagues (item 6 in Table 4.22). Our data revealed that criteria of the rubric for the speaking test were those (item 8 in Part 2 from students' questionnaire and item 12 in Part 2 from teachers' questionnaire): clarity of expression, content, organization, vocabulary, accuracy, fluency, pronunciation, body language, eye contact, students' enthusiasm. In addition, teachers should take their students' opinions while designing a rubric. For example, some suggestions of the students for a rubric in the interviews were as follows:

*"I wish I could omit grammar as a criterion in the rubric" (Audio Recording: S-95/high level).*

*"Sometimes, I cannot form sentences accurately or I can confuse word order. I think my teacher should not care them too much" (Audio Recording: S-106/high level).*

*"I wish I had been assessed in accordance with criteria like fluency and pronunciation" (Audio Recording: S-77/midlevel).*

*"My teacher should assess speaking skill in regard to vocabulary knowledge rather than grammar" (Audio Recording: S-72/high level).*

Table 4.22

*Descriptive Statistics for Rating Scales from Teachers' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
5. I prepare a rubric or a checklist for the speaking test.	22	4.41	1.00	4.5	0	9.1	22.7	63.6
25. I used a checklist or a rubric to assess their performance.	22	4.41	.79	0	4.5	4.5	36.4	54.5
32. I gave each student's mark during the speaking test.	22	3.95	1.25	4.5	13.6	9.1	27.3	45.5
6. My colleagues and I agreed on rubric items together.	22	3.82	1.25	4.5	13.6	18.2	22.7	40.9
33. I gave each student's mark at the end of the speaking test after all students had finished it.	22	2.91	1.57	27.3	18.2	13.6	18.2	22.7

Teachers (90.9%) declared that they used a checklist or a rubric to assess students' performance (item 25 in Table 4.22). In reference to item 11 in Figure 4.4, 22.7% of the teachers used analytic scale for its reliability whereas 40.9% of the teachers preferred holistic scales because of its easiness. There is discrepancy between quantity of item 11 in Figure 4.4 and item 25 in Table 4.22 because 31.8% of the teachers abstained from answering item 11 in Figure 4.4. Nonetheless, most teachers implied that they used a checklist or a rubric in order to increase test reliability.

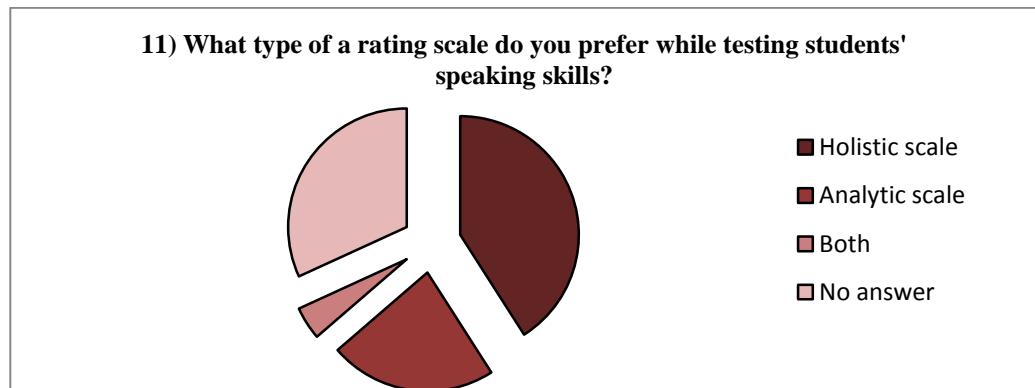


Figure 4.4. The demonstration of the teachers' preferences for the rating scales

In addition, 56.2% of the students remarked that their teachers assessed their performance on the basis of a rubric or a checklist (item 32 in Table 4.23). Maybe, the others (43.8%) did not witness scoring process, and they did not see any rubric in front of their teachers. Inasmuch as, 40.9% of the teachers notified that they gave each student's mark at the end of the speaking test after all students had finished it (item 33 in Table 4.22) while 72.8% of the teachers narrated that they gave each student's mark during the speaking test (item 32 in Table 4.22). Even so, unfortunately, some teachers ignored

importance of the rubric, and they did not use it. However, rubric is a sort of step to ensure reliability of the test.

Table 4.23

*Descriptive Statistics for Rating Scales from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
32. My teachers assessed my performance on the basis of a rubric or a checklist.	358	3.54	1.34	10.6	13.7	18.2	27.7	28.5

**4.1.11. Feedback**

50.8% of the students notified that their grades they got from speaking test correctly reflected their speaking ability (item 33 in Table 4.24). In other words, nearly one-half of the students were satisfied with their teachers' feedback in terms of their score. 42.7% of the students also mentioned that their teachers gave them feedback verbally related to their performance in the speaking test (item 17 in Table 4.24). These students illustrated their feedback in the interviews:

*"My teacher reported that I got low mark because I did not revise vocabularies so much"* (Audio Recording: S-29/not happy with his/her mark).

*"My teacher articulated that I mispronounced vocabularies on account of my stress"* (Audio Recording: S-50/not happy with his/her mark).

*"My teacher signified that I should study harder. S/he led me to work efficiently"* (Audio Recording: S-44/not happy with his/her mark).

Table 4.24

*Descriptive Statistics for Feedback from Students' Perspective*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
33. The grade I got from speaking test correctly reflected my speaking ability.	358	3.38	1.29	11.7	12.8	23.2	31.8	19
17. My teachers gave me feedback related to my performance in the speaking test.	358	3.13	1.46	18.2	17.3	20.4	24	18.7

Feedback should be meaningful or sightful for the students. In the contrary case, they were confused as in the following example:

*My teacher says that I have to study harder. I know it but I do not know how to study efficiently. He or she further expresses that I need to memorize vocabularies and grammar rules. I have already made these practices; even so, I cannot get high mark. I cannot pronounce well. My teacher never speaks in Turkish during class hours. We are not ready for it. We, except a few students, cannot understand our teacher. As a result, we cannot learn or speak anything. Our teacher knows it but still speaks English during class hours* (Audio Recording: S-58/not happy with his/her mark).

So, the teachers need to be clear and comprehensible as giving feedback. However, some students had preconceived notions about how their teachers would handle with request of the feedback, and these preconceived notions could change with respect to pleasure of the students' marks. Students' expressions from the interviews were as follows:

*“Even if I opposed to my mark, probably it would not change, and my teacher would inform about issues I should pay attention to for next time” (Audio Recording: S-7/not happy with his/her mark).*

*“Even if I talked about my mark, nothing would change, and probably my teacher would tell me to study harder and revise more often” (Audio Recording: S-89/partly happy with his/her mark).*

*“If I were not happy with my mark, I would talk about reason for my mark with my teacher. My teacher could assess me again or explain reasons for my mark” (Audio Recording: S-1/happy with his/her mark).*

*“My teacher pays attention to why I object to my mark. If I am wrong, s/he explains my weaknesses or mistakes. In the contrary case, if I am right, my teacher gives me extra points to make up his/her mistake” (Audio Recording: S-25/happy with his/her mark).*

Whatever the circumstances are, students are in need of receiving meaningful feedback to identify their level, improve their weaknesses, and enrich their strengths. Feedback can be in the form of score or verbal descriptors of the students' performance. According to Bachman & Palmer (1996), feedback is very effective in developing positive attitude towards tests on the behalf of the students. So teachers should be diligent in giving feedback.

## **4.2. Students' Perception and Attitudes**

On the purpose of conceiving students' attitudes towards speaking tests, those titles are handled with: experience of the students, speaking tests, test tasks, class activities, feelings of the students and feedback. The items inquiring into students' perception and attitudes regarding the speaking test in Anatolian high schools and their quantitative analysis are presented in the following tables.

### **4.2.1. Experience of the Students**

Students (87.2%) did not have any experience regarding speaking tests in advance (item 1 in Figure 4.5).



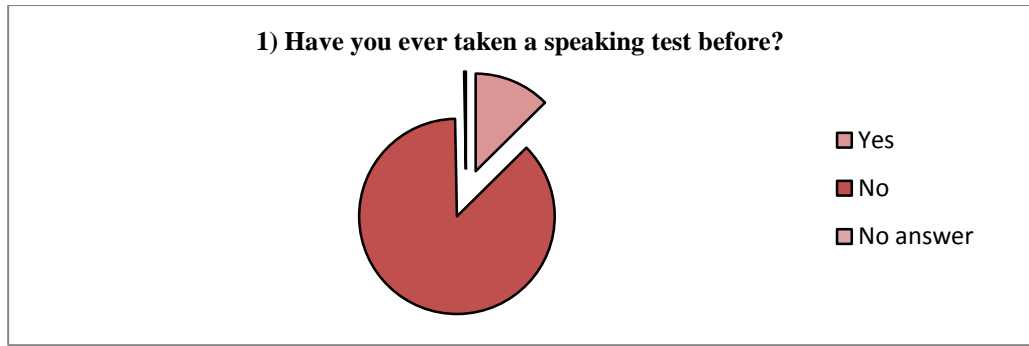


Figure 4.5. The percentages of the students who experience speaking tests

Even though some students admitted that they took speaking tests in private schools or language courses previously, 87.2% of the students did not take a speaking test before (item 1 in Figure 4.5). Inasmuch as, speaking tests were not compulsory in primary schools. As a result, 42.8% of the students acknowledged that they were also inexperienced with how they should get ready for the speaking test (item 5 in Table 4.25). Namely, they did not know the procedure of the test, rubric, and interpretation of score, which could affect test reliability (Henning, 1987). So the students needed much more guidance of their teachers.

Table 4.25

*Descriptive Statistics for Students' Experience*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
5. I was inexperienced with how I should get ready for speaking test.	358	3.15	1.42	15.9	21.2	18.2	23.2	19.6

**4.2.2. Speaking Tests**

Our data revealed that students' attitude towards speaking tests was generally positive. Students (63.4%) advocated that speaking test was useful to improve their English (item 37 in Table 4.26). Besides, 31.9% of the students implied that if they did not have to take a speaking test, they would not spend so much time to improve their speaking skill (item 39 in Table 4.26). In other words, students focused on the speaking skill more by way of speaking tests as washback. This idea was reiterated in the interviews as follows:

*“My fluency, pronunciation and self-expression skill have progressed. My self-confidence has also been reformed after speaking test” (Audio Recording: S-61/high level).*

*“I have improved myself in speaking. My self-confidence is getting increase, and I can speak more easily” (Audio Recording: S-26/high level).*

*“The speaking test encouraged me to study English. It contributed to my vocabulary knowledge. When I researched on a topic and talked about it, I learnt new vocabularies which were related to the topic” (Audio Recording: S-1/high level).*

Table 4.26

*Descriptive Statistics for Students’ Perception and Attitudes Regarding the Speaking Test*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
37. Speaking test is useful to improve my English.	358	3.74	1.23	8.1	7.8	19.8	31.3	32.1
39. If I did not have to take a speaking test, I would not spend so much time to improve my speaking skill.	358	2.87	1.36	19.6	23.2	24.6	16.8	15.1
44. It is waste of time to assess speaking skill.	358	2.16	1.36	43.6	26	13.7	3.9	12.8

On the other hand, 16.7% of the students stood against speaking tests (item 44 in Table 4.26), and they characterized speaking tests as waste of time. In addition, the students narrated it with the following statements in the interviews:

*“We have already been doing pencil-and-paper tests. I do not know why my teacher does speaking tests. It seems useless” (Audio Recording: S-16/low level).*

*“We do not speak English in our daily life. It is not necessary to be tested for speaking skill” (Audio Recording: S-85/low level).*

*“Speaking tests are pointless. They are useful just for memorization because I have to memorize what I will say before a speaking test” (Audio Recording: S-12/low level).*

According to Bachman & Palmer (1996), students’ attitudes have the potential impact on their performance. If students adopt negative attitude, their achievement levels will be generally low as can be seen above statements of the students in the interviews.

#### 4.2.3. Test Tasks

Students affirmed that the tasks and activities were comprehensible (62.3%) and from daily life (63.1%) in the speaking test (item 19 and 21 in Table 4.27). Moreover, 70.7% of the students did not embrace that some tasks and activities were non-class and unexpected activities (item 22 in Table 4.27). As an inference, test tasks were reasonable for students.

Table 4.27

*Descriptive Statistics for Students' Perception and Attitudes Regarding Test Tasks*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
19. The tasks and activities were loud and clear in the speaking test.	358	3.76	1.26	6.4	7.3	22.9	33.5	28.8
21. The tasks and activities were from daily life in the speaking test.	358	3.76	1.22	6.1	8.9	20.7	33.2	29.9
14. It is hard to express myself clearly in the speaking test.	358	2.71	1.25	18.2	29.3	28.5	11.5	12.3
42. Speaking test is the most difficult test.	358	2.48	1.39	30.4	29.1	18.4	6.4	15.4
23. The tasks and activities were difficult.	358	2.33	1.14	27.4	32.4	27.1	6.7	6.1
22. Some tasks and activities were non-class and unexpected activities in the speaking test.	358	2.18	1.26	36.6	34.1	14	6.7	8.1

In respect to difficulty, 59.8% of the students notified that the tasks and activities were difficult (item 23 in Table 4.27). Besides, the students (59.5%) contradicted that the speaking test was the most difficult test (item 42 in Table 4.27). Nonetheless, 23.8% of the students agreed that it was hard for them to express themselves clearly in the speaking test (item 14 in Table 4.27) much as 47.5% of them objected it. They explained the case in the interviews:

*"We have difficulty in pronunciations" (Audio Recording: S-42/midlevel).*

*"I can select wrong word, and I cannot form sentences accurately" (Audio Recording: S-69/midlevel).*

*"I forgot what I would say during the speaking test while I could speak better" (Audio Recording: S-55/high level).*

What is more, some students had multiple difficulties in speaking test like: understanding questions / activities, answering appropriately to questions, first thinking in Turkish then translating to English as speaking, speaking fluently and accurately, finding correct vocabulary as speaking, pronouncing correctly, contextualizing, and overcoming the stress (item 7 in Part 2 from students' questionnaire). All these difficulties are natural for the students. The difficulties will decrease as long as the students develop their English level, and hopefully, they will get used to taking such tests.

#### 4.2.4. Class Activities

According to the students (53.9%), the speaking test was parallel with class activities (item 20 in Table 4.28). Therefore, the students (47.4%) sustained that studying for the speaking test made contribution to their class performances (item 38 in Table 4.28).

Even 15.1% of the students (item 36 in Table 4.28) indicated that if they did not have to take a speaking test, they would not take part in speaking activities in the lessons. Namely, speaking tests enhanced value of the speaking skill and activities for the students. They also articulated their opinions in the interviews:

*“... Thanks to studying for the speaking test, both I grasped the lesson better, and got a high mark” (Audio Recording: S-107/high level).*

*“After the speaking test, I studied harder to speak more fluently and accurately. I also tried to participate in speaking activities more often in the classroom” (Audio Recording: S-103/high level).*

Table 4.28

*Descriptive Statistics for Students' Perception and Attitudes Regarding Class Activities*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
20. The speaking test was parallel with class activities.	358	3.53	1.29	10.6	9.5	24.9	27.4	26.5
38. I can also use many of the things in lessons which I have studied for the speaking test.	358	3.41	1.29	10.9	12	27.9	25.1	22.3
25. I wanted to do more speaking activities in class after this test.	358	2.63	1.38	29.3	19.6	22.9	15.9	11.7
24. Speaking activities during lessons were not sufficient.	358	2.39	1.34	31.6	29.9	19.8	8.4	7.8
36. If I did not have to take a speaking test, I would not take part in speaking activities in the lessons.	358	2.34	1.27	31.6	29.6	22.6	7	8.1

Students (61.5%) submitted that speaking activities during lessons were sufficient (item 24 in Table 4.28). Yet, 27.6% of the students presented that they wanted to do more speaking activities in class after the speaking test (item 25 in Table 4.28). Insofar as students gained awareness of the speaking skill as washback of the speaking test. Much though most of the students were satisfied with class activities, some students upheld that class activities were inadequate, and they also depicted it in the interviews as follows:

*“Although we never speak English in the classroom, I am supposed to speak English during the test. I want to change class activities because class activities and test tasks are not parallel” (Audio Recording: S-86/high level).*

*“I think speaking test is useful, but our class activities are crap... English, in games and films, is more useful” (Audio Recording: S-88/midlevel).*

*“We do not do speaking activities in class hours. Only my teacher speaks English, and we just listen to him/her. Mostly, we do not understand him/her (during the lesson). Actually, two or three students from our class can understand my teacher (the others cannot understand the teacher)” (Audio Recording: S-58/low level).*

#### 4.2.5. Feelings of the Students

Each student is unique, and each one has very different feelings. So it is challenging to generalize them. Nonetheless, I will try to clarify the students’ feelings. According to data retrieved from the questionnaires, students (60.6%) opposed that they found themselves the most unsuccessful in speaking tests (item 43 in Table 4.29). Nevertheless, 31.3% of the students reported that their personal characteristics such as coyness, timidity, nervousness, etc. stopped them from speaking English during the test (item 15 in Table 4.29). A student uttered his/her stress with the following statement in the interview:

*“I feel very stressful (during the test). Even if I want to speak, I cannot” (Audio Recording: S-105/midlevel).*

Table 4.29

*Descriptive Statistics for Students’ Perception and Attitudes Regarding Students’ Feelings*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
41. Speaking test makes me nervous.	358	3.16	1.45	18.4	15.6	23.7	16.2	25.1
29. I performed better than I expected in speaking test.	358	2.91	1.28	17	21.2	29.6	18.2	13.4
15. My personal characteristics such as coyness, timidity, nervousness, etc. stop me from speaking English during the test.	358	2.87	1.34	19	23.5	26	15.4	15.9
40. I think I need to take speaking tests more frequently.	358	2.77	1.44	24.3	21.2	24.9	14.5	14.8
35. I feel more relaxed in speaking tests in contrast with pen-and-paper tests.	358	2.73	1.43	27.1	20.7	21.5	14.5	15.4
43. Speaking test is a test that I find myself the most unsuccessful.	358	2.42	1.39	32.1	28.5	19.6	7	12.3

Moreover, 41.3% of the students signified that speaking test made them nervous (item 41 in Table 4.29). There might be many reasons for this, and some of them were illustrated in the interviews as follows:

*“It was the first time we took a speaking test so I felt nervous” (Audio Recording: S-48/high level).*

*“I got nervous about my performance: whether I would pronounce clearly or not, whether I would forget what I would say or not...” (Audio Recording: S-61/high level)*

*“I did not want to take low mark so I was stressed” (Audio Recording: S-28/midlevel).*

However, 31.6% of the students remarked that they performed better than they expected in speaking test (item 29 in Table 4.29). At the same time, 29.9% of the students reflected that they felt more relaxed in speaking tests in contrast with pencil-and-paper tests (item 35 in Table 4.29). These students verbalized their opinions as follows:

*“... There is just one correct answer in pencil-and-paper tests, but it is not so in speaking tests” (Audio Recording: S-2/high level).*

*“Speaking tests show us that English does not consist of just pencil-and-paper tests. It is also for communication” (Audio Recording: S-62/midlevel).*

*“Speaking tests are more advantageous than pencil-and-paper tests. I learn new vocabularies and their pronunciations. I upgrade my speaking skill. I observe my strength and weakness in speaking. I cannot do them in pencil-and-paper tests” (Audio Recording: S-41/high level).*

As a whole, the students were not demoralized. They believed in themselves, and they were motivated. Even so, students were not so eager to be tested more frequently. 45.5% of the students (item 40 in Table 4.29) antagonized that they needed to take speaking tests more frequently, and they agreed that it was satisfactory.

#### **4.2.6. Feedback**

Language tests are applied in order to diagnose students' level, their strengths and weaknesses. In our study, students (58.4%) also believed that speaking test was essential to identify their speaking level (item 34 in Table 4.30). Whereas 37.4% of the students (item 30 in Table 4.30) pointed out that they noticed their strengths, 49.2% of the students (item 27 in Table 4.30) represented that they noticed their weaknesses in speaking after the test.

Table 4.30

#### *Descriptive Statistics for Students' Perception and Attitudes Regarding Feedback*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
34. Speaking test is essential to identify my speaking level.	358	3.56	1.29	11.2	8.9	20.9	30.7	27.7
27. I notice my weaknesses in speaking after speaking test.	358	3.39	1.24	10.6	11.5	27.9	29.1	20.1
30. I notice my strengths in speaking after speaking test.	358	3.08	1.30	14.8	18.2	28.8	21.5	15.9
28. I feel that I need to study English harder after speaking test.	358	2.94	1.31	17.6	20.7	26	21.8	13.4

Feedback raises consciousness among students and teachers. As such, teachers can design their teaching program depending upon students' levels. Besides, students can plan their studying and review their progress. To illustrate, though the percentage was low, 35.2% of the students (item 28 in Table 4.30) acknowledged that they needed to study English harder after the speaking test. In short, speaking tests were motivator for those students. Some expressions from the interviews also supported this idea:

*"I study on my weakness, and improve my English. I revise English more at home"* (Audio Recording: S-24/high level).

*"We see our weaknesses, and we comprehend what we need to study"* (Audio Recording: S-108/midlevel).

*"Speaking tests help me to improve English and reiterate it"* (Audio Recording: S-43/high level).

### 4.3. Differences between Male and Female Students

When independent sample t-tests were computed for the gender difference, the data revealed that there was no statistically significant difference between male and female students in regard to their perception and attitudes towards the speaking test ( $t=-.630$ ,  $p>.05$ ) as illustrated in Table 4.31.

Table 4.31

*Attitude of Male and Female Students towards Speaking Tests*

Gender	N	Mean	Sd	t	df	Sig. (2-tailed)
Female	222	2.93	.285	-.630	356	.529
Male	136	2.96	.405			

### 4.4. Differences among Schools in Terms of Students' Placement Scores

One-way ANOVA was computed to compare the students' perception and attitudes towards the speaking test among Anatolian high schools in terms of the students' placement scores. Although there were six different Anatolian high schools in the study, they were divided into three groups in terms of the students' placement scores in TEOG exam, which was a test for the eight grade students for their transition from primary to high schools. The schools were classified as high level (two of them), middle level (two of them), and low level (two of them). According to the school categories, descriptive results of one-way ANOVA are shown in Table 4.32.

Table 4.32

*Descriptive Results of One-Way ANOVA in Terms of School Categories*

School Categories	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
High level	119	3.1352	.26814	.02458	3.0865	3.1839	2.52	4.55
Medium level	114	3.2125	.36457	.03414	3.1449	3.2802	2.00	4.77
Low level	125	3.1393	.28395	.02540	3.0890	3.1895	2.30	3.70
<b>Total</b>	358	3.1612	.30834	.01630	3.1292	3.1933	2.00	4.77

The results of the one-way ANOVA indicated that there was no statistically significant difference among schools with regard to the students' perception and attitudes towards the speaking test ( $p > .098$ ) as presented in Table 4.33.

Table 4.33

*Attitude of the Schools towards Speaking Tests*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.441	2	.220	2.335	.098
Within Groups	33.501	355	.094		
<b>Total</b>	33.942	357			

#### 4.5. Teachers' Perception and Attitudes

Teachers' attitudes towards speaking tests are reviewed under those titles: experience of the teachers, speaking tests, test tasks, class activities, feelings of the teachers, rating scales, and scoring. The items searching teachers' perception and attitudes regarding the speaking test in Anatolian high schools and their quantitative analysis are presented in the following tables.

##### 4.5.1. Experience of the Teachers

Only 13.6% of the teachers set forth that they were inexperienced with speaking tests (item 3 in Table 4.34). All teachers (item 1 and 2 in Table 4.34) specified that they prepared and administered a speaking test before. According to the teachers (77.3%), it was not difficult to prepare the speaking test (item 45 in Table 4.34). While 31.8% of the teachers notified that they had some worries about the administration of a speaking test (item 11 in Table 4.34) before, 45.4% of them did not agree.



Table 4.34

*Descriptive Statistics for Teachers' Experience*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
2. I have administered a speaking test before.	22	4.86	.35	0	0	0	13.6	86.4
1. I have prepared a speaking test before.	22	4.82	.39	0	0	0	18.2	81.8
52. All teachers should be trained through in-service training on how speaking skill can be tested.	22	4.09	.97	0	9.1	13.6	36.4	40.9
11. I have some worries about the administration of a speaking test before.	22	2.73	1.31	22.7	22.7	22.7	22.7	9.1
45. It is difficult for me to prepare the speaking test.	22	2.00	1.02	36.4	40.9	9.1	13.6	0
3. I am inexperienced with speaking tests.	22	1.95	1.13	40.9	40.9	4.5	9.1	4.5

Besides, the teachers (90.9%) reflected that they did not take part in interlocutor training sessions (item 1 in Figure 4.6) before speaking tests. Purely 9.1% of the teachers claimed that they took part in interlocutor training sessions in universities in Figure 4.6.



Figure 4.6. The percentage of the teachers who do interlocutor training

Ministry of National Education did not offer teachers such an opportunity for training. However, teachers (77.3%) relied that all teachers should be trained through in-service training on how speaking skill can be tested (item 52 in Table 4.34).

#### 4.5.2. Speaking Tests

Teachers (95.4%) pointed out that they knew speaking tests were compulsory in high schools (item 38 in Table 4.35), and they generally adopted positive attitude towards speaking tests. However, some teachers did not assess speaking skills of the students by alleging difficulties of testing. According to these teachers, difficulties of the speaking tests were physical factors (classroom, organization, time, noise), validity and reliability problems, objective scoring, time-consuming, tiring, and need for expert and experienced teachers (item 13 in Part 2 from teachers' questionnaire). Moreover, the teachers considered that students did not feel volunteer to improve their speaking skills due to

reality of university exam, which ignored speaking skill. Students focused on items in the university exam, and as a washback, they neglected the speaking skill, which is not part of the university exam. Basturkmen (2001) also sets forth that university entrance exam focuses on testing reading skills, knowledge of vocabulary and grammar, and translation by ignoring writing, speaking and listening skills in Turkey.

Nonetheless, the teachers attached importance to speaking tests, and they (77.3%) relied that the speaking test was necessary to see the students' speaking level (item 40 in Table 4.35). even a teacher pleaded it with the following statement in the interview:

*"It is necessary to test students' speaking ability because we cannot evaluate it during class hours"* (Audio Recording: T-8).

What is more, the teachers (81.8%) noted that the speaking test helped the students to notice their weaknesses in their speaking performances (item 42 in Table 4.35). Accordingly, the teachers could plan their class activities by considering the students' weaknesses and needs.

Table 4.35

*Descriptive Statistics for Teachers' Perception and Attitudes Regarding the Speaking Test*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
38. I know that the speaking test is compulsory at high schools.	22	4.59	.59	0	0	4.5	31.8	63.6
41. The speaking test encourages students to use their speaking ability.	22	4.27	1.16	0	9.1	13.6	31.8	31.8
42. The speaking test helps the students to notice the weaknesses in their speaking performances.	22	4.14	.71	0	0	18.2	50	31.8
40. The speaking test is necessary to see the students' speaking level.	22	4.09	.86	0	4.5	18.2	40.9	36.4
43. Trying to test speaking skill is a waste of time.	22	1.95	1.21	54.5	13.6	13.6	18.2	0

The teachers (68.1%) repugated that trying to test speaking skill was a waste of time (item 43 in Table 4.35). As a washback effect, they (63.6%) believed that the speaking test encouraged students to use their speaking ability (item 41 in Table 4.35). Some expressions from the teachers' interviews were as follows:

*"Students concentrate more on speaking activities in class when they are assessed in speaking"* (Audio Recording: T-12).

*"Speaking tests ensure that students practice English speaking. Thus, they can express themselves better. In addition, students enhance their vocabulary knowledge, and give examples from their personal experiences"* (Audio Recording: T-2).

### 4.5.3. Test Tasks

All teachers asserted that the tasks and activities were comprehensible and from daily life (item 29 and 30 in Table 4.36). Furthermore, teachers (95.5%) considered that the tasks and activities were parallel with those in the course book (item 28 in Table 4.36). Teachers (95.4%) detected that the tasks and activities were not above the students' abilities and levels (item 27 in Table 4.36).

Table 4.36

#### *Descriptive Statistics for Teachers' Perception and Attitudes Regarding Test Tasks*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
29. The tasks and activities were clear and comprehensible.	22	4.82	.39	0	0	0	18.2	81.8
28. The tasks and activities were parallel with those in the course book.	22	4.73	.55	0	0	4.5	18.2	77.3
30. The tasks and activities were from daily life.	22	4.64	.49	0	0	0	36.4	63.6
27. The tasks and activities were not above the students' abilities and levels.	22	4.59	.90	4.5	0	0	22.7	72.7

Furthermore, item 9 in Part 2 from teachers' questionnaire tried to elicit teachers' preferences related to test tasks. Lack of agreeable answer, teachers preferred those test tasks in speaking tests: reading aloud, sentence repetition/transformation/construction/completion, completing a dialogue, translating/interpreting a text/dialogue, information-gap activities, talking about pictures/video/film, question and answer, giving instructions/explanation/ description, problem solving activity, role-playing, verbal essay, oral interview, discussion, oral presentation, and writing a short story/paragraph or completing a dialogue.

### 4.5.4. Class Activities

Teachers (72.8%) did not approve that if speaking skills were not tested, they would not spend so much time on teaching the speaking skill (item 44 in Table 4.37).

Table 4.37

#### *Descriptive Statistics for Teachers' Perception and Attitudes Regarding Class Activities*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
44. If speaking skills were not tested, I would not spend so much time on teaching the speaking skill.	22	1.95	1.09	45.5	27.3	13.6	13.6	0

It could be deduced that teachers placed importance on the speaking skill, and they employed speaking activities during class hours. Even they (54.5%) believed that they should assess speaking skills by continuous assessment (item 14 in Table 38).

Table 4.38

*Descriptive Statistics for Teacher Preferences for the Frequency of Speaking Tests*

<b>Item 14: How should speaking skills be tested?</b>	<b>Frequency</b>	<b>Percent</b>
a) By continuous assessment (all the time in class).	12	54.5
b) By achievement tests every month.	1	4.5
c) By an achievement test at the end of each term.	3	13.6
d) By using all the suggestions mentioned above.	3	13.6
No answer	3	13.6
<b>Total</b>	<b>22</b>	<b>100</b>

Notwithstanding, some teachers depicted that they did not spend time on speaking activities during class hours. For example;

*“We have to assess speaking skill although we do not make any speaking activities in classroom. Besides, we do not speak English continuously during class hours. So, we need extra class hours for speaking activities” (Audio Recording: T-11).*

*“There should be extra class hours for speaking skill. Classes should not be so crowded because it is getting more difficult for me to make some activities in these (crowded) classes” (Audio Recording: T-5).*

Teachers need to admit that assessment does not only imply assigning grades to students, but it is also a fundamental key to regulate the process of teaching and learning. Speaking tests will be fairer if students experience speaking activities during class hours. Otherwise, students will do speaking tests without practicing their speaking skills.

#### **4.5.5. Feelings of the Teachers**

Teachers (68.2%) refuted that they felt more anxious and nervous during the speaking test in contrast with pencil-and-paper tests (item 20 in Table 4.39). Namely, most of the teachers felt relaxed. In addition, teachers (68.2%) contradicted that it was stressful to administer the speaking test for them (item 46 in Table 4.39). However, some teachers uttered contrary sentences in the interviews as follows:

*“Crowded classes complicate speaking tests. It was really weary so it would be better to have fewer students in classes” (Audio Recording: T-1).*

*“Some students do not want to speak in spite of all my effort” (Audio Recording: T-5).*

Table 4.39

*Descriptive Statistics for Teachers' Feelings*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
21. The students' anxiety level is very high during the speaking test.	22	3.82	1.14	4.5	4.5	31.8	22.7	36.4
31. Some students whose English level is high cannot perform well due to their test anxiety during the speaking test.	22	2.82	1.00	9.1	31.8	27.3	31.8	0
46. It is stressful to administer the speaking test for me.	22	2.23	1.15	31.8	36.4	9.1	22.7	0
20. I feel more anxious and nervous during the speaking test in contrast with pen-and-paper tests.	22	2.00	1.15	45.5	22.7	22.7	4.5	4.5

Furthermore, 59.1% of the teachers supposed that the students' anxiety level was very high during the speaking test (item 21 in Table 4.39). 31.8% of the teachers pleaded that some students whose English level was high could not perform well due to their test anxiety during the speaking test (item 31 in Table 4.39). Teachers also delineated their students' anxiety levels in the interviews as follows:

*"Their classmates may humiliate test takers during their performance"* (Audio Recording: T-9).

*"Students do not want to speak. They feel ashamed, and they cannot express themselves freely. They forget what they will say due to their stress"* (Audio Recording: T-13).

*"Some students have self-confidence problems. They are afraid of mispronunciation. When they take a negative response from their classmates, they feel disappointment"* (Audio Recording: T-2).

#### 4.5.6. Rating Scales

Teachers generally adopted a positive attitude towards rubrics. Even 90.9% of the teachers antagonized that rubric was useless in the speaking test (item 49 in Table 4.40). They (81.8%) thought that the rubric for the speaking test was satisfactory to assess the students effectively (item 35 in Table 4.40). In addition, the teachers (90.9%) embraced that rubric items for the speaking test were clear (item 34 in Table 4.40).

Table 4.40

*Descriptive Statistics for Teachers' Perception and Attitudes Regarding Rating Scales*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
34. Rubric items for the speaking test were clear.	22	4.59	.66	0	0	4.5	36.4	54.5
35. The rubric for the speaking test was satisfactory to assess them effectively.	22	4.23	.75	0	0	18.2	40.9	40.9
47. All teachers should use the rubric developed and imposed by the Ministry of	22	3.32	1.17	0	31.8	27.3	18.2	22.7

National Education.								
The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
48. It is difficult to prepare rubric for the speaking test.	22	2.68	1.17	18.2	27.3	27.3	22.7	4.5
49. Rubric is useless in the speaking test.	22	1.64	.79	50	40.9	4.5	4.5	0

Only 27.2% of the teachers (item 48 in Table 4.40) accepted that it was difficult to prepare a rubric for the speaking test. As an implication, most of them could design a rubric or get a ready-made rubric at ease. Besides, Regulation for Secondary Education Institutions (2014) suggested the CEFR as the base. In my interviews, teachers were not aware of it, and they just preferred getting rubric on the Internet. However, 40.9% of the teachers (item 47 in Table 4.40) relied that all teachers should use the rubric developed and imposed by the Ministry of National Education.

#### 4.5.7. Scoring

Teachers (59.1%) justified that speaking skills could be assessed accurately (item 39 in Table 4.41). 40.9% of the teachers agreed (item 15 in Table 4.41) that it was better to assign only one interlocutor and one assessor during speaking tests. 40.9% of the teachers also partly agreed with it. However, all of them assessed their students alone, and they did not cooperate with their colleagues.

Table 4.41

#### *Descriptive Statistics for Teachers' Perception and Attitudes Regarding Scoring*

The items	N	Mean	Sd	S.D. %	D. %	P.A. %	A. %	S.A. %
39. Speaking skills can be assessed accurately.	22	3.82	1.00	0	9.1	31.8	27.3	31.8
37. I think I am a reliable rater.	22	4.50	.59	0	0	4.5	40.9	54.5
51. I believe that all the teachers at my school are objective.	22	3.55	1.18	4.5	18.2	13.6	50	9.1
15. It is better to assign only one interlocutor and one assessor as graders.	22	3.55	1.29	9.1	4.5	40.9	18.2	22.7
50. It is difficult to be objective as marking the speaking test for me.	22	1.86	.77	31.8	54.5	9.1	4.5	0

In this sense, self-confidence of the teachers was high, and they (95.4%) thought they were a reliable rater (item 37 in Table 4.41). Only 4.5% of the teachers affirmed that it was difficult to be objective as marking the speaking test (item 50 in Table 4.41). All the same, the teachers (59.1%) believed that all teachers at their school were objective (item 51 in Table 4.41).

## **CHAPTER V: CONCLUSION**

Speaking tests as the outcome of innovations and reforms in English language curriculum have been compulsory in high schools in Turkey since 2014 (Regulation for Secondary Education Institutions, 2014). The aim of this study was to identify how English language teachers assess the 9<sup>th</sup> grade students' speaking skills in Anatolian high schools. Another aim of the study was to investigate the students' and the teachers' perceptions and attitudes toward speaking tests in these high schools. Moreover, the study intended to research whether students' attitudes differ by their gender and schools.

The research was designed as the convergent parallel design in mixed method. Questionnaires and interviews for both teachers and students were the main instruments to obtain data. Questionnaires were administered to 358 students from the CEFR A1 & A2 level (basic user) and 22 English language teachers who were teaching those students as participants. In addition, 112 students and 13 teachers were subsequently interviewed via audio recordings in order to enrich quantitative data. The participants were the 9<sup>th</sup> grade students and their teachers from six different Anatolian high schools which assessed students' speaking skills in Denizli. The 9<sup>th</sup> grade students were selected since they had six hours of English classes per week while the other graders had four hours of English classes. All in all, the whole process of speaking tests in Anatolian high schools, students' and their teachers' attitudes emerged from the quantitative and qualitative data. In the analysis of the data, Statistical Package for Social Sciences (SPSS) 22.0 was used for questionnaires. What is more, content analysis was applied for the interviews after they were transcribed.

In this chapter, the findings of the study will be reviewed and discussed. Then, foregrounding implications of the study will be included. Lastly, suggestions for further research will be provided.

### **5.1. Discussion**

Major findings of the study are discussed in parallel with research questions.

#### **5.1.1. Research Question 1: How are Speaking Skills Assessed in Anatolian High Schools?**

The results of the quantitative and qualitative data were examined, and it was found out that though English language curriculum recommended the CEFR as basis, there was no standard for speaking tests in Anatolian high schools. This result is in contrast with the study of Alastrué & Pérez-Llantada (2010) which was followed through oral English test as a part of the new model of University entrance exam. According to Regulation for

Secondary Education Institutions (2014), teachers just had to do speaking test at least once in each academic term in our study. In Sook's (2003) study, Korean Junior Secondary School English teachers conducted speaking assessments at least once a year in their classrooms. According to Lee (2010), English teachers held speaking assessment once or twice in a semester in secondary schools in South Korea.

Our data highlighted that teachers designed test tasks, setting, materials and rubric previously on their own. Accordingly, teachers informed students about testing and scoring procedure of speaking tests. Students also confirmed their teachers about informing them with regard to the speaking test; but, as a matter of course, there were some students who had contrary opinions. As a consequence, students planned their studying depending upon this information, and mostly preferred way of studying for the students was to cram for the test individually.

Regarding the time, teachers had to determine and announce the date of the speaking test at the beginning of each academic term in our study (Regulation for Secondary Education Institutions, 2014). The teachers tested the students during English lessons, and they generally allocated minimum 2 minutes and maximum 15 minutes for each student. Both teachers and students were generally satisfied with timing despite some students' complaint about it.

Teachers designed the speaking test under the influence of the curriculum, English course book, students' level and interest, daily social issues related to students' ages and current interests, and affairs in students' real life. Even though the students did various tasks, presentation was the most favorite task in our study in contrast with the following studies: (a) rote memorization of text dialogues was a common practice for speaking assessment in Korea (Nagata, 1995), (b) interview was viewed as a more real type of communication in Alastrué & Pérez-Llantada's (2010) study, (c) conversation was the most favored one in the study of Restrepo et al. (2003), (d) the most preferred speaking activity was asking and answering questions in Güllüoğlu's (2004) study. Whereas N. Underhill (1987) claims that discussion is "most natural" of task types (p. 45), A. Hughes (2003) alleges that interviews are "the most common format for the testing" (p. 119).

Testing environment was a classroom in our study. While a teacher was assessing a student, the other students were also in the classroom. As a result, some teachers and students were not pleased with this testing condition because of noise, stress and distractibility. Besides, students were assessed individually. The teachers did not prefer pair or group test tasks because they thought that pair and group tasks could create



administration problems. However, the students preferred paired tasks like role-plays, information gap, etc. so as not to feel alone and stressful so much in Güllüoğlu's (2004) study which revealed students' attitudes towards speaking tests at Gazi University Preparatory School of English.

According to the teachers and the students in our study, the materials used in speaking tests were those: course book, pictures, objects, realia, films, video, EBA software, interactive board, question pool, topics for discussion and presentation, news, magazines, brochures, dialogue for role play, incomplete stories, reading passages, dialogues, computer, and mobile phone for recording. It is significant to select intelligible and clear materials for teachers. Even Fulcher and Davidson (2007) advocate that test materials should be standardized for reliability and validity. S.K. Kitao & K. Kitao (1996) further express that teachers can control vocabulary and grammatical structures required through careful selection of the material.

There was only one teacher during speaking tests in our study. Accordingly, a teacher adopted both interlocutor's and assessor's role. Yet, some teachers reported that they adopted merely assessor's role. This was because the teachers selected presentation task as an assessment tool, and they considered that there was no need for interaction. While the students were presenting, teachers just listened to and assess them. It was recommended to have an assessor and interlocutor during the test for a reliable assessment in Sook's (2003) and Höl's (2010) studies. Furthermore, Bachman and Palmer (1996) suggest two or three assessors ideally. N. Underhill (1987) also claims that "the more assessors you have for any single test... the more reliable the score will be" (p. 89).

Teachers encouraged and motivated the students to speak more often, do more class activities, and improve their English levels. Moreover, teachers supported their students with their positive behavior during speaking tests in our study. We can see similar results in Höl's (2010) study that interlocutors or assessors encouraged and motivated students during speaking test. However, students were not so eager for speaking tests firstly, yet they adopted positive attitude over time in Güllüoğlu's (2004) study. On the one hand, teachers tried to motivate students but these students disregarded speaking tests in studies of Alastrué & Pérez-Llantada (2010) and Majid, Eng & Samad (2007). On the other hand, students remarked that they needed more support from teachers in Majid, Samad, Muhamad and Vethamani's (2011) study.

Our study noted that the teachers tolerated their students' mistakes, and gave high marks so as to show their support. But this support changed up to the teachers' feelings. Therefore, it created reliability problems, and it was also evidence of objectivity problems.

However, the teachers were not aware of these problems, and even they stated that they were objective in assessing the students' performances in the questionnaires. Furthermore, the students acknowledged that their teachers were objective in speaking tests, and some students also notified that their grade they got from the speaking test accurately reflected their speaking ability. In other words, nearly one-half of the students were satisfied with the teachers' feedback in our study. In Güllüoğlu's (2004) study, teachers supposed that they scored their students' performance as objectively as possible. However, teachers did not believe that speaking assessment was objective whereas students believed that they were assessed objectively in Höl's (2010) study.

According to our results, there was no standard rubric for speaking tests in high schools though rubric was compulsory. Teachers declared that they used a checklist or a rubric to assess students' performance, and the students approved them. Teachers could design it on the basis of the CEFR, or get a readymade rubric. Chuang (2009) suggests rating scales to assess Taiwanese EFL learners' oral proficiency and explain those reasons for the rating scales: (a) teachers can follow the scales/criteria/standards, (b) teachers can maintain the intra-rater reliability and validity of the test, (c) teachers can give feedback to students based on the descriptor of the rating scales.

Criteria of rubric for the speaking test in our study were those: clarity of expression, content, organization, vocabulary, accuracy, fluency, pronunciation, body language, eye contact, and students' enthusiasm. Restrepo et al. (2003) advocated that there should be similar criteria to assess students' speaking skills more systematically and objectively, and designated those criteria in their study: grammar, pronunciation, conveying meaning, fluency, vocabulary and comprehension and competence. On the other hand, Sinwongsuwat (2012) highlights that "rubric should be adjusted to accommodate features of naturally occurring conversation, rather than simply focusing on discrete items such as pronunciation, fluency, grammar, vocabulary, and comprehension as traditionally practiced" (p. 81). In conclusion, setting criteria is indeed a crucial and difficult issue in assessment since each teacher has his/her own way of perceiving students' performance no matter how clear and refined the criteria are (Restrepo et al., 2003). Notwithstanding, the aim of the criteria is to minimize measurement errors and standardize scoring process.

### **5.1.2. Research Question 2: What are Students' Perception and Attitudes towards the Speaking Test?**

Most students did not have experience in advance because speaking tests were not compulsory in primary schools. The students' speaking skills were generally assessed first

in high schools. The attitude of the students towards speaking tests was generally positive. Our results were parallel with Güllüoğlu's (2004) and Duran's (2011) studies whose participants also adopted positive attitudes towards speaking tests.

On the one hand, the students advocated that speaking test was useful to improve their English in our study. Moreover, students believed that speaking test was essential to identify their speaking level. The participants in Höl's (2010) study pointed out that it was essential to have a speaking test to identify their strengths and weaknesses in speaking English. On the other hand, 16.7% of the students stood against speaking tests in our study. Even they characterized speaking tests as waste of time. Furthermore, students were very resistant towards English and were not so motivated to speak in Majid, Eng & Samad's (2007) study which analyzed the effects of the school-based oral English test (OET) on teacher perceptions and practices in Malaysia. In the same way, teachers complained about students' lack of interest in Alastrué & Pérez-Llantada's (2010) study.

Our results revealed that test tasks were reasonable for the students. They notified that the tasks and activities were not so difficult although 23.8% of the students had difficulty in expressing themselves clearly in the speaking test. The students also affirmed that the tasks and activities were comprehensible and from daily life. Students shared similar opinions in Höl's (2010) study, and they also stated that test tasks were comprehensible and from daily life.

Students submitted that speaking activities during lessons were sufficient in our study. Even some students presented that they wanted to do more speaking activities in class after the speaking test. Insofar as these students gained awareness of speaking skill as washback of speaking tests. On the contrary, some students mentioned that they could not practice their speaking skills enough in class hours due to the lack of emphasis on speaking in Güllüoğlu's (2004) study.

Students had different feelings regarding speaking tests in our study. 31.3% of the students reported that their personal characteristics such as coyness, timidity, nervousness, etc. stopped them from speaking English during the test. In addition, 41.3% of the students signified that speaking test made them nervous. At the same time, 29.9% of the students reflected that they felt more relaxed in speaking tests in contrast with pencil-and-paper tests. As a whole, students were not demoralized. They believed in themselves, and they were motivated. Even so, students were not so eager to be tested more frequently, and some of them antagonized that they needed to take speaking tests more frequently. Besides, students from Höl's (2010) study also did not want to have speaking tests more

often though they knew the importance of the speaking skill. Höl (2010), who sought students' and instructors' attitudes towards the speaking test at a School of Foreign Languages in Pamukkale University, explained its reason for students' high anxiety and stress.

### **5.1.3. Sub-Question of Research Question 2: Are There any Differences between Male and Female Students?**

Our data revealed that there was no statistically significant difference between male and female students regarding their attitudes towards speaking tests ( $t=-.630, p>.05$ ).

### **5.1.4. Sub-Question of Research Question 2: Are There any Differences among the Schools in Terms of Students' Placement Scores?**

Our data indicated that there was no statistically significant difference among the schools with regard to their students' perception and attitudes towards speaking tests ( $p>.098$ ).

### **5.1.5. Research Question 3: What are the Teachers' Perception and Attitudes towards the Speaking Test?**

All teachers specified that they prepared and administered a speaking test before so they were experienced teachers in our study. Teachers also pointed out that they knew speaking test was compulsory in high schools, and they generally adopted positive attitudes towards speaking tests. Moreover, they believed that speaking tests were necessary to see students' speaking levels, and the tests encouraged students to perform their speaking ability. On the one hand, there are many researches which support teachers' positive attitude towards speaking test such as Sook's (2003), Güllüoğlu's (2004), Alastrué & Pérez-Llantada (2010), and Duran's (2011) studies. On the other hand, Korean teachers asserted that speaking tests did not assess students' authentic communicative competence (Sook, 2003). Additionally, most teachers had pessimistic attitudes towards speaking assessment in Lee's (2010) research because they characterized assessment methods as a lack of spontaneous responses and interpersonal exchanges

Teachers did not have in-service training relevant to speaking assessment (constructing, administering or rating process) in our study. What is more, Ministry of National Education did not offer teachers such an opportunity. However, teachers believed that they needed to be trained through in-service training on how speaking skill can be tested. There are similar samples: Korean Junior Secondary School English teachers denied having sufficient opportunities for in-service education in Sook's (2003) study. Restrepo et

al. (2003) underlined the need of in-service training courses in the area of assessment for the teachers. Furthermore, Lee (2010) suggested teacher training, cooperation with an English native teacher, and downsizing the number of students per class to improve speaking assessment in secondary schools in South Korea.

Our results indicated that some teachers did not assess their students' speaking skills by alleging difficulties of testing. According to the teachers, difficulties of the speaking tests were physical factors (classroom, organization, time and noise), validity and reliability problems, objective scoring, time-consuming, tiring, and need for expert and experienced teachers. Korean teachers also identified those constraints in Sook's (2003) study: large classes, excessive work in addition to classroom teaching, lack of effective and efficient assessment instruments, difficulty in eliciting students' responses, difficulty in ensuring reliability, teachers' low English proficiency, students' low English proficiency, and few opportunities for retraining. In Lee's (2010) study, the most challenging factor was an oversized class problem. As such, more than 36 students in a class were too large a number to complete assessment within a class period of 45 minutes.

According to our data, the teachers preferred multiple test tasks in speaking tests. All teachers asserted that the tasks and activities were comprehensible, from daily life and parallel with students' course books. The teachers also detected that the tasks and activities were not above the students' abilities and levels. Furthermore, the teachers refuted that they felt more anxious and nervous during the speaking test in contrast with pencil-and-paper tests. Insofar as some teachers employed speaking activities during class hours even though some of them did not, and thus they felt more relaxed.

Teachers adopted positive attitudes towards rubrics in our study. As such, they antagonized that rubric was useless in the speaking test. The teachers' self-confidence was high, and they thought they were reliable raters. Therefore, some teachers justified that speaking skills could be assessed accurately in our study. What is more, participants believed that their performance was assessed accurately in Duran's (2011) study. Teachers also advocated that speaking assessment should be a reliable and objective measurement tool in Lee's (2010) study. However, Korean Junior Secondary School English teachers had little confidence in ensuring reliability of scoring, and they did not feel ready enough to construct and administer communicative speaking assessment in Sook's (2003) study because they were not equipped with an adequate theory of speaking assessment.

## 5.2. Pedagogical Implications

Speaking tests have been compulsory in high schools since 2014 in Turkey (Regulation for Secondary Education Institutions, 2014). However, there has not been any standard for speaking tests in high schools yet. The study is one of the pioneering studies searching speaking tests because there seems no prominent research reflecting how speaking tests are applied in high schools in Turkey or what their teachers and students attitudes are although there have been few studies like Güllüoğlu (2004), Höl, (2010), Duran (2011) on perception and attitudes of teachers and students in universities. Therefore, the study might contribute to improve high school programs in terms of teaching and testing speaking to some extent.

The results of the study may provide valuable indications to policy makers who take decisions about English teaching and testing system. They can select new textbooks and new teaching methods, which focus on speaking skills more. They may standardize speaking test for reliability and validity, hold out inservice training for teachers, and organize follow-up seminars in high schools. In addition, they may embark on an enterprise so as to manage difficulties during speaking tests.

The findings of the study may be a trigger for other teachers who deal with similar problems concerning speaking test. They can compare their own way with the findings, and they can keep going on what they do or they can make some changes. Next, this study may guide teachers to get awareness about the attitudes and perceptions of their students towards testing speaking, and they may try to find out alternative ways to help their students. Consequently, feelings of the teachers and students can be interpreted better by way of the study.

## 5.3. Suggestions

This study raises a number of opportunities for future research. For example, replication of this study will be more beneficial if further research includes more students with different levels of English and different age range. The students in this study were the 9<sup>th</sup> graders, and they were the CEFR A1 & A2 level according to English language curriculum (Ministry of National Education, 2014). Speaking performances from different levels of English might have a different effect on the results. Besides, more teachers might participate in further study because there were only 22 teachers in this study. In future research, the number of the schools can be extended because there were just six Anatolian high schools in this study, and all of them were from Denizli. Further study may be conducted with different type of high school, not only Anatolian high schools around the country.

In a further study, teachers can record speaking performance of the students during tests so as to investigate how speaking assessment is performed in a more detailed way. Researchers can also analyze rating scales to identify teachers' scoring way. Moreover, further research may also take a historical perspective and ask if speaking tests have significantly changed in the last several decades, including a significant shift in test tasks and attitude. Comparative questions can also be asked with regard to differences in assessing the speaking skill between Turkey and other countries. In conclusion, more research will in fact be necessary to refine and further elaborate our findings.

## REFERENCES

- Alastrué, R.P. & Pérez-Llantada, C. (2010). A longitudinal attitude survey on English oral skills: classroom, curriculum, learning and pedagogy implications. *Revista de Humanidades*, 16, 309-326.
- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.
- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Aljohani, M. (2017). Principles of constructivism in foreign language teaching. *Journal of Literature and Art Studies*, 7(1), 97-107. doi: 10.17265/2159-5836/2017.01.013.
- Allan, D. (1999). Testing and assessment. *English Teaching Professional*, 11, 19-20.
- Aloni, N. (1997). A redefinition of liberal and humanistic education. *International Review of Education*, 43, 87-107.
- Arslan, R.Ş. (2013). An investigation of prospective English language teachers' spoken communication skills: a case from Turkey. *Hacettepe University Journal of Education*. 28(1), 27-40.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, 17(1), 1-42. Retrieved from <http://dx.doi.org/10.1177/026553220001700101>.
- Bachman, L.F. & Palmer, A.S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Bagarić, V. & Djigunović, J.M. (2007). Defining communicative competence. *Metodika*, 8, 94-103.
- Bamgbose, A. (2003). A recurring decimal: English in language policy and planning. *World Englishes*, 22(4), 419-431.
- Basturkmen, H. (2001). Descriptions of spoken language for higher level learners. *ELT Journal*, 55(1), 4-12.
- Bayyurt, Y. (2012). Proposing a model for English language education in the Turkish socio-cultural context. In Y. Bayyurt and Y. Bektaş-Çetinkaya (Eds.), *Research Perspectives on Teaching and Learning English: Policies and Practices* (299-310). Frankfurt am Main: Peter Lang.
- Berry, V. (1993). Personality characteristics as a potential source of language test bias. In A. Huhta, K. Sajavaara and S. Takala (Eds.), *Language testing: new openings* (115-124). Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Berry, V. (2007). *Personality differences and oral test performance*. Frankfurt: Peter Lang.



- Best, J.W. and Kahn, J.V. (1998). *Research in Education* (8th Ed.). Boston: Allyn and Bacon.
- Basic National Education Law [Milli Eğitim Temel Kanunu]. (2012). T.C. Resmi Gazete [T.R. Official Gazette], 28261, 11 April 2012.
- Boyacıoğlu, F. (2015). The historical development of the foreign language education in Ottoman Empire. *Procedia - Social and Behavioral Sciences*, 174, 651-657.
- Breen, M. (1987). Learner contributions to task design. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (23-46). Englewood Cliffs, NJ: Prentice-Hall.
- Brereton, J.L. (1944). *The case for examinations: an account of their place in education with some proposals for their reform*. Cambridge: Cambridge University Press.
- Brindley, G. (1991). Defining language ability: the criteria for criteria. In S. Anivan (Ed.), *Current developments in language testing* (139-164). Singapore: Regional Language Centre.
- British Council, (1983). *VOTE: Oral testing*. London: British Council English Languages Services Department and the Design, Production and Publishing Department.
- Brooks, J.G. and Brooks, M.G., (1993). *In search of understanding: the case for constructivist classrooms*. Alexandria, Virginia, USA: Association for Supervision and Curriculum Development.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: co-constructing a better performance. *Language Testing*, 26(3), 341-366. Retrieved from <http://dx.doi.org/10.1177/0265532209104666>.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20, 1-25.
- Brown, A. (2005). *Interviewer variability in oral proficiency interviews*. Frankfurt: Peter Lang.
- Brown, A., McNamara, T., Iwashita, N. and O'Hagan, S. (2001). *Investigating raters' orientations in specific-purpose task-based oral assessment. (TOEFL: research and development project reports)*. Princeton, NJ: Educational Testing Service.
- Brown, H.D. (1994). *Principles of language learning and teaching*. Englewood Cliffs, NJ: Prentice Hall.
- Brown, H.D. (2004). *Language assessment. Principles and classroom practices*. NY, USA: Pearson Education.
- Brown, J.D. (1996). *Testing in language programs*. Upper Saddle River, New Jersey: Prentice Hall Regents.
- Brown, K. (2002). Ideology and context: world Englishes and EFL teacher training. *World Englishes*, 21(3), 445-448.
- Bygate, M. (Ed.) (1987). *Speaking*. Oxford: Oxford University Press.

- Bygate, M., Skehan, P. & Swain, M. (Eds.) (2001). *Researching pedagogic tasks: second language learning, teaching, and testing*. London: Longman.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication* (2-27). London: Longman.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carey, L.M. (1988). *Measuring and evaluating school learning*. Newton, Massachusetts: Allyn and Bacon, Inc.
- Carroll, B.J. (1980). *Testing communicative performance*. London: Pergamon.
- Carroll, J.B. (1961). *Fundamental considerations in testing for English language proficiency of foreign students*. New York: McGraw-Hill.
- Cheng, L. & Curtis, A. (2004). Washback or backwash: a review of the impact of testing on teaching and learning. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: research contexts and methods* (3-17). Mahwah, New Jersey: Lawrence Erlbaum.
- Chuang, Y.Y. (2009). Foreign language speaking assessment: Taiwanese college English teachers' scoring performance in the holistic and analytic rating methods. *Asian EFL Journal*, 11(1), 150-173.
- Clahsen, H. (1984). The acquisition of German word order: a test case for cognitive approaches to second language acquisition. R. Andersen, *Second languages: a cross-linguistic perspective* (219-242). Rowley, MA: Newbury House.
- Clark, J.L.D. (1972). *Foreign language testing: theory and practice*. Philadelphia: Center for Curriculum Development.
- Cohen, A.D. (1994). *Assessing language ability in the classroom (teaching methods)*. Boston: Heinle ELT.
- Cohen, A.S. and Wollack, J.A. (2006). Test administration, security, scoring and reporting. In R.L. Brennan (Ed.), *Educational Measurement* (Forth edition) (355–386). New York: American Council on Education/Praeger.
- Coombe, C., Folse, N. and Hubley, N. (2007). *A practical guide to assessing English language learners*. USA: The University of Michigan Press.
- Cooper, P.A. (1993). Paradigm shifts in designed instruction: from behaviorism to cognitivism to constructivism. *Educational Technology*, 33(5), 12-19.
- Corder, S.P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council of Europe.

- Creswell, J.W. & Clark, V.L.P. (2011). *Designing and conducting mixed methods research* (second edition). California, America: Sage Publications.
- Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.
- Çaykan, I. (2001). *Material development for testing oral skills communicatively for intermediate students*. (Unpublished master's thesis). Hacettepe University, Ankara, Turkey.
- Çetintaş, B. & Genç A. (2001). Eğitim reformu sonrası Anadolu liselerinde yabancı dil öğretimi [Foreign language teaching at Anatolian high schools following the education reform]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 51-56.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- De Matos, F.G. (2005). The EFL teacher as a humanizer. *International Association of Teachers of English as a Foreign Language (IATEFL) Issues*, Issue 182, 7. Retrieved from <https://associates.iatefl.org/pages/materials/voicespdf/pd15.pdf>.
- Demircan, Ö. (1988). *Dünden bugüne Türkiye'de yabancı dil* [Foreign language in Turkey from past to today]. İstanbul, Turkey: Remzi.
- Demirel, Ö. (1990). *Yabancı dil öğretimi. İlkeler, yöntemler, teknikler* [Foreign language teaching. Principles, methods, techniques]. Ankara, Turkey: USEM.
- Dewey, J. (1938). *Experience and education*. New York: Macmillan/Collier.
- Douglas, D. (1997). TOEFL monograph series: testing speaking ability in academic contexts: theoretical considerations (Number 8). Princeton, New Jersey: Educational Testing Service.
- Douglas, D. (Eds.) (2014). *Understanding language testing* (Second Edition). New York, USA: Routledge.
- Driscoll, M.P. (2000). *Psychology of learning for instruction*. Needham Heights, MA: Allyn & Bacon.
- Ducasse, A.M. & Brown, A. (2009). Assessing paired orals: raters' orientation to interaction. *Language Testing*, 26(3), 423-443. Retrieved from <http://dx.doi.org/10.1177/0265532209104669>.
- Duran, Ö. (2011). *Teachers' and students' perceptions about classroom-based speaking tests and their washback*. (Unpublished master's thesis). Bilkent University, Ankara, Turkey.
- East, M. (2008). *Dictionary use in foreign language writing exams: impact and implications*. Amsterdam, Netherlands/Philadelphia, PA: John Benjamins.
- East, M. (2012). *Task-based language teaching from the teachers' perspective: insights from New Zealand*. Amsterdam, Netherlands / Philadelphia, PA: John Benjamins. Retrieved from <http://dx.doi.org/10.1075/tblt.3>.

- East, M. (2015). Taking communication to task again: What difference does a decade make? *The Language Learning Journal*, 43(1), 6–19. Retrieved from <http://dx.doi.org/10.1080/09571736.2012.723729>.
- East, M. (2016). Assessing foreign language students' spoken proficiency: stakeholder perspectives on assessment innovation. Singapore: Springer.
- Education First English Proficiency Index, (2017). The world's largest ranking of countries by English skills (Seventh Edition). Retrieved from <https://www.ef.com.tr/epi/regions/europe/turkey/>.
- Egyud, G. & Glover, P. (2001). Readers respond: Oral testing in pairs-secondary school perspective. *ELT journal*, 55(1), 70-76.
- Elias, J.L. & Merriam, S. (1980). Philosophical foundations of adult education. Malabar, FL: Krieger.
- Ellis, R. (2008). The study of second language acquisition. Oxford: Oxford University Press.
- Encyclopedia Britannica. (1993). (15. Edition). (Volume 2). (234-235). İstanbul: Ana Yayıncılık.
- Ergün, M. (2000). Medreseden mektebe Osmanlı eğitim sistemindeki değişme [Transformation from madarsas to modern school in Ottoman education system]. *Yeni Türkiye*, 32,2000, 735-753.
- Foot, M.C. (1999). Relaxing in pairs. *ELT Journal*, 53(1), 36-41.
- Foreign Language Teaching and Learning Act [Yabancı Dil Eğitimi ve Öğretimi Kanunu]. (1983). T.C. Resmi Gazete [T.R. Official Gazette], 18196, 19 October 1983.
- Friedman, J. (1994). Cultural identity and global process. London: Sage.
- Fulcher, G. (1996). Testing tasks: issues in task design and the group oral. *Language Testing*, 13(1), 23-51. Retrieved from <http://dx.doi.org/10.1177/026553229601300103>.
- Fulcher, G. (1997). The testing of speaking in a second language. In C. Clapham and D. Corson (Eds.), *Encyclopedia of language and education: language testing and assessment* (vol. 7, 75-85). Amsterdam: Kluwer Academic Publishers.
- Fulcher, G. (2010). Practical language testing. London: Hodder Education.
- Fulcher, G. (Ed.) (2014). Testing second language speaking (Second Edition). London & New York: Routledge.
- Fulcher, G. & Davidson, F. (2007). Language testing and assessment: an advanced resource book. New York: Routledge.
- Gadamer, H.G. (1975). Truth and method. New York: Crossroad.

- Galaczi, E.D. (2010). Paired speaking tests: an approach grounded in theory and practice. In J. Mader & Z. Ürkün (Eds.), *Recent approaches to teaching and assessing speaking*. (International Association of Teachers of English as a Foreign Language (IATEFL) TEA SIG Conference proceedings). Canterbury, England: IATEFL Publications.
- Grabe, W. (1988). English, information access, and technology transfer: a rationale for English as an international language. *World Englishes*, 7(1), 63-72.
- Güllüoğlu, Ö. (2004). Attitudes towards testing speaking at Gazi University preparatory school of English and suggested speaking tests. (Unpublished master's thesis). Gazi University, Ankara, Turkey.
- Haines, M.M., Stansfeld, S.A., Head, J. and Job, R.F.S. (2002). Multilevel modelling of aircraft noise on performance tests in schools around Heathrow Airport London. *Journal of Epidemiology and Community Health*, 56, 139-144.
- Hall, G. (2011). *Exploring English language teaching: language in action*. New York: Routledge.
- Hambleton, R.K. (1978). On the use of cut off scores with criterion-referenced tests in instructional settings. *Journal of Educational Measurement*, 15, 277-290.
- Harmer, J. (2003). *The practice of English language teaching*. New York: Pearson Education.
- Harris, D.P. (1969). *Testing English as a second language*. New York: McGraw Hill Company.
- Heaton, J.B. (Eds.) (1988). *Writing English language tests (Second Edition)*. New York: Longman.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford, England: Oxford University Press.
- Henning, G. (1987). *A Guide to language testing: development, evaluation and research*. Rowley, Massachusetts: Newbury House.
- Herman, J.L., Aschbacher, P.R. and Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Holt, D.G. & Willard-Holt, C. (2000). Let's get real: students solving authentic corporate problems. *Phi Delta Kappan*, 82(3), 243-246.
- Hosseini, A.S. & Azarnoosh, M. (2014). Iranian EFL instructor's oral assessment practices: purposes, methods, procedures. *Procedia - Social and Behavioral Sciences*, 98, 653-658.
- Höl, D. (2010). Perceptions and attitudes of the English language instructors and preparatory students towards testing speaking communicatively in the school of foreign languages at Pamukkale University. (Unpublished master's thesis). Pamukkale University, Denizli, Turkey.

- Hsieh, H.F. & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Huerta-Macias, A. (1995). Alternative assessment: responses to commonly asked questions. *TESOL Journal*, 5(1), 8-11.
- Hughes, A. (1983). Second language learning and communicative language teaching. In K.A. Johnson and D.A. Porter (Eds.), *Perspectives in communicative language teaching*. New York: Academic Press.
- Hughes, A. (2003). *Testing for language teachers* (Second edition). Cambridge, UK: Cambridge University Press.
- Hughes, R. (2002). *Teaching and researching speaking*. London: Longman
- Hymes, D.H. (1972). On communicative competence. In J.B. Pride & J. Holmes (Eds.), *Sociolinguistics* (269-293). Baltimore, USA: Penguin Education.
- Ingram, E. (1977). Basic concepts in testing. In J.P.B. Allen and A. Davies (Eds.), *Testing and experimental methods* (Volume 4). Oxford: Oxford University Press.
- Işık, A. (2011). Language education and ELT materials in Turkey from the path dependence perspective. *Hacettepe University Journal of Education*, 40, 256-266.
- Jones, R.L. (1977). Testing: a vital connection. In J.K. Philips (Ed.), *The language connection: from the classroom to the world* (237-265). Skokie, IL: National Textbook Company.
- Jones, R.L. and Spolsky, B. (Ed.) (1975). *Testing language proficiency*. Arlington, Virginia: Centre for Applied Linguistics.
- Kachru, B.B. (1992). *The other tongue* (Second Ed.). Urbana, IL: University of Illinois Press.
- Khatib, M., Sarem, S.N. & Hamidi H. (2013). Humanistic education: concerns, implications and applications. *Journal of Language Teaching and Research*, 4(1), 45-51.
- Kırkgöz, Y. (2005a). English language teaching in Turkey: challenges for the 21st century. In G. Braine (Ed.), *Teaching English to the world: history, curriculum, and practice* (159-75). Mahwah: Lawrence Erlbaum Associates.
- Kırkgöz, Y. (2005b). Motivation and student perception of studying in an English-medium university. *Journal of Language and Linguistic Studies*, 1(1), 101-123.
- Kırkgöz, Y. (2007). English language teaching in Turkey: policy changes and their implementations. *RELC Journal*, 38(2), 216-228.
- Kırkgöz, Y. (2009). Globalization and English language policy in Turkey. *Educational Policy*, 23, 663-684.
- Kızıltepe, Z. (2000). Attitudes and motivation of Turkish EFL students towards second language learning. *ITL: Review of Applied Linguistics*, 129(1), 141-168.

- Kitao, S.K. & Kitao, K. (1996). Testing speaking. *ERIC Document Reproduction*, Service No. ED 398261, 1-7.
- Kitao, S.K. & Kitao, K. (1998). Testing communicative competence. *The Internet TESL Journal*, 2, 1-4.
- König, G.Ç. (1990). The place of English in Turkey. In D. Bozer (Ed.), *The birth and growth of a department: department of English language and literature: 25th Anniversary (157-167)*. Ankara, Turkey: Hacettepe University Publications.
- Lazaraton, A. (2001). Teaching oral skills. In Marianne Celce-Murcia (Ed.), *Teaching English as a second foreign language* (103-115). Boston: Heinle and Heinle.
- Lee, S. (2010). Current practice of classroom speaking assessment in secondary schools in South Korea. (Unpublished master's thesis). The University of Queensland, Australia.
- Long, M.H. (1985). A role for instruction in second language acquisition: task-based language training. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (77-99). Clevedon, UK: Multilingual Matters.
- Lozovska-Güneş, A. (2010). Differences and challenges involved in the assessment of speaking skill: the case of three universities in Ankara. (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.
- Luoma, S. (Ed.) (2004). *Assessing speaking*. Cambridge, England: Cambridge University Press.
- Madsen, H.S. (1983). *Techniques in testing*. Oxford: Oxford University Press.
- Majid, Z.A., Eng, M.L.L. & Samad A.A. (2007). The school based oral English test (OET): perceptions and practices of selected MARA Junior Science College Teachers. *Jurnal Pendidikan*, 2(2), 1-18. Retrieved from [https://works.bepress.com/arshad\\_abdsamad/16/](https://works.bepress.com/arshad_abdsamad/16/).
- Majid, Z.B.A., Samad, A.B.A., Muhamad, M.B. and Vethamani, M.E. (2011). The school-based oral English test: similarities and differences in opinion between teachers and students. *The English Teacher*, 40, 113-128.
- Majidifard, E., Shomoossi, N. & Ghourchaei, B. (2014). Risk taking, gender and oral narrative proficiency in Persian. *Procedia - Social and Behavioral Sciences*, 98, 1085-1092.
- Marlowe, B.A. & Page, M.L. (2005). *Creating and sustaining the constructivist classroom* (Second Ed.). Thousand Oaks, CA: Corwin Press.
- Maslow, A.H. (1962). Notes on being-psychology. *Journal of Humanistic Psychology*, 2(2), 47-71. doi:10.1177/002216786200200205
- May, L. (2009). Co-constructed interaction in a paired speaking test: the rater's perspective. *Language Testing*, 26(3), 397-422. Retrieved from <http://dx.doi.org/10.1177/0265532209104668>.

- McEwen, N. (1995). Educational accountability in Alberta. *Canadian Journal of Education*, 20, 27-44.
- McKenna, G. (1995). Learning theories made easy: humanism. *Nursing Standard*, 2(9), 29-31.
- McNamara, T. (1996). *Measuring second language performance*. Chicago: Addison Wesley Longman.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- McNamara, T. & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics*, 18(1), 89-95.
- Medgyes, P. (1986). Queries from a communicative teacher. *English Language Teaching Journal*, 40(2). Retrieved from <https://doi.org/10.1093/elt/40.2.107>.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83-108.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Miles, M. & Huberman, A.M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage Publications.
- Ministry of National Education. (2013). İlköğretim İngilizce dersi (2, 3, 4, 5, 6, 7 ve 8. sınıflar) öğretim program [Primary education English language curriculum (the 2<sup>nd</sup>-8<sup>th</sup> grades)]. Retrieved from <http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327>.
- Ministry of National Education. (2014). Ortaöğretim İngilizce dersi (9, 10, 11 ve 12. sınıflar) öğretim program [Secondary education English language curriculum (the 9<sup>th</sup>-12<sup>th</sup> grades)]. Retrieved from <http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=342>.
- Ministry of National Education - Board of Education. (2014). Ortaöğretim kurumları haftalık ders çizelgeleri [Weekly course schedule of secondary education institutions (the 9<sup>th</sup>-12<sup>th</sup> grades)]. Retrieved from <http://ttkb.meb.gov.tr/www/haftalik-ders-cizelgeleri/kategori/7>.
- Ministry of National Education - Board of Education. (2017). İlköğretim kurumları haftalık ders çizelgesi [Weekly course schedule of primary education institutions (the 2<sup>nd</sup>-8<sup>th</sup> grades)]. Retrieved from <http://ttkb.meb.gov.tr/www/haftalik-ders-cizelgeleri/kategori/7>.
- Ministry of National Education - Board of Education. (2018). Ortaöğretim kurumları haftalık ders çizelgeleri [Weekly course schedule of secondary education institutions (the 9<sup>th</sup>-12<sup>th</sup> grades)]. Retrieved from <http://ttkb.meb.gov.tr/www/haftalik-ders-cizelgeleri/kategori/7>.
- Mishra, C. (2000). Humanistic approach to education. *Journal of NELTA*, 5(2), 26-29.



- Morrow, K. (1977). Techniques of evaluation for a notional syllabus. London: Royal Society of Arts.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing* (29-42). Windsor, Berks: NFER-Nelson.
- Morrow, K. (2012). Communicative language testing. In C. Coombe, P. Davidson, B. O'Sullivan and S. Stoyhoff (Eds), *The Cambridge guide to second language assessment* (140-146). Cambridge: Cambridge University Press.
- Morse, J.M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40, 120-123.
- Moskowitz, G. (1978). Caring and sharing in the foreign language class: a sourcebook on humanistic techniques. Massachusetts: Newbury House.
- Mousavi, S.A. (2002). An encyclopedic dictionary of language testing (Third Edition). Taiwan: Tung Hua Book Company.
- Nagata, H. (1995). Testing oral ability: ILR and ACTFL oral proficiency interviews. *JALT Applied Materials*, 12, 108-118.
- Norris, J.M. (2009). Task-based teaching and testing. In M.H. Long & C.J. Doughty (Eds.), *The handbook of language teaching* (578-594). Oxford, UK: Wiley-Blackwell.
- Nunan, D. (1991). Communicative tasks and the language curriculum. *Tesol Quarterly*, 25(2), 279-295.
- Nunan, D. (2004). Task-based language teaching. Cambridge, England: Cambridge University Press. Retrieved from <http://dx.doi.org/10.1017/cbo9780511667336>.
- Ockey, G.J. (2001). Is the oral interview superior to the group oral? *Working Papers, International University of Japan*, 11, 22-40.
- Oller, J.W. (1976). Language testing. In R. Wardhaugh and H.D. Brown (Eds.), *A survey of applied linguistics* (275-300). Ann Arbor: University of Michigan Press.
- O'Malley, M. & Valdez-Pierce, L. (1996). Authentic assessment for English language learners: practical approaches for teachers. New York: Addison Wesley.
- O'Sullivan, B. (2000). Towards a model of performance in oral language testing. (Unpublished doctoral dissertation). University of Reading, Berkshire, England.
- O'Sullivan, B. (2006). Modelling performance in oral language tests: language testing and evaluation. Frankfurt: Peter Lang.
- O'Sullivan, B. (2008). Notes on assessing speaking. [Online]. Retrieved from <http://lrc.cornell.edu/events/past/2008-2009/papers08/osull1.pdf> (2 November 2017).
- O'Sullivan, B. (2013). Assessing spoken language: scoring validity. In D. Tsagari, S. Papadima-Sophocleous, S. Ioannou-Georgiou (Eds.), *Language Testing and Evaluation: international experiences in language testing and assessment* (Volume 28, 261-274). Frankfurt: Peter Lang Edition.

- O'Sullivan, B. and Weir, C. (2002). Research issues in testing spoken language. (Unpublished research report). Cambridge: University of Cambridge Local Examinations Syndicate.
- Önal, A. (2010). Testing and assessment of speaking skills in preparatory classes. (Unpublished master's thesis). Selçuk University, Konya, Turkey.
- Özbay, F. (n.d.). Türkiye'de yabancı dil öğretiminin tarihi geçmişi ve Türkiye'nin günümüzdeki yabancı dil politikası [Historical background of teaching foreign language in Turkey and Turkey's foreign language policy today]. Retrieved from [https://www.google.com.tr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjH143mxbPVAhXJshQKHbBwD-cQFggnMAA&url=http%3A%2F%2Fwww.onlinearabic.net%2Fsizden\\_gelenler\\_dosya%2F8043444.doc&usg=AFQjCNEirXuFln3GKq6Hi76tzeb2P3xFuQ](https://www.google.com.tr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjH143mxbPVAhXJshQKHbBwD-cQFggnMAA&url=http%3A%2F%2Fwww.onlinearabic.net%2Fsizden_gelenler_dosya%2F8043444.doc&usg=AFQjCNEirXuFln3GKq6Hi76tzeb2P3xFuQ) (18 July 2017).
- Paker, T. (2012). Türkiye'de neden yabancı dil (İngilizce) öğretemiyoruz ve neden öğrencilerimiz iletişim kurabilecek düzeyde İngilizce öğrenemiyor? [Why do not we teach foreign language (English), and why cannot our students learn English at a certain level to communicate?]. *Pamukkale University Journal of Education*, 32(2), 89-94.
- Paker, T. (2013). The backwash effect of the test items in the achievement exams in preparatory classes. *Procedia - Social and Behavioral Sciences*, 70, 1463-1471.
- Patterson, C.H. (1973). Humanistic education. Englewood Cliffs, NJ: Prentice-Hall.
- Piaget, J. (1955). The construction of reality in the child. Retrieved from [www.marxists.org/reference/subject/philosophy/works/fr/piaget2.htm](http://www.marxists.org/reference/subject/philosophy/works/fr/piaget2.htm).
- Pienemann, M. (1998). Language processing and second language development: processability theory. Amsterdam: John Benjamins.
- Popham, W.J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.
- Powers, D.E., Albertson, W., Florek, T., Johnson, K., Malak, J., Nemceff, B., Porzuc, M., Silvester, D., Wang, M., Weston, R., Winner, E. and Zelazny, A. (2002). Influence of irrelevant speech on standardized test performance (TOEFL Research Report-68). Princeton, NJ: Educational Testing Service.
- Prabhu, N.S. (1987). Second language pedagogy. Oxford: Oxford University Press.
- Rami, J., Lorenzi, F. & Lalor, J. (2009, September). The application of constructivist assessment practices in a teacher-training programme: a tool for developing professional competencies. Bera Annual Conference, University of Manchester, Manchester.
- Regulation for Higher Education Institutions Which Teach a Foreign Language and Teach through a Foreign Language [Yükseköğretim Kurumlarında Yabancı Dil Eğitim-Öğretimi ve Yabancı Dille Eğitim-Öğretim Yapan Yükseköğretim Kurumları Hakkında Yönetmelik]. (1984). T.C. Resmi Gazete [T.R. Official Gazette], 18378, 20 April 1984. Retrieved from <http://www.resmigazete.gov.tr/default.aspx>.

- Regulation for Secondary Education Institutions [Ortaöğretim Kurumları Yönetmeliği]. (2014). T.C. Resmi Gazete [T.R. Official Gazette], 29118, 13 September 2014.
- Regulation for Secondary Education Institutions [Ortaöğretim Kurumları Yönetmeliği]. (2016). T.C. Resmi Gazete [T.R. Official Gazette], 29871, 28 October 2016.
- Restrepo, A.P.M., Aristizábal, L.D., Orozco, F.C., Monsalve, S.G., Orozco, L.A.L. and Urán, M.P. (2003). Assessing spoken language in EFL: beliefs and practices. *Revista Universidad EAFIT*, [S.l.], 39(129), 63-73.
- Reves, T. (1991). From testing research to educational policy: a comprehensive test of oral proficiency. In J.C. Alderson and B. North (Eds), *Language testing in the 1990s* (178-188). London: Modern English Publications and the British Council.
- Richards, J.C. & Rodgers, T.S. (2014). *Approaches and methods in language teaching* (Third Ed.). Cambridge, England: Cambridge University Press.
- Roach, J.O. (1945). Some problems of oral examinations in modern languages: an experimental approach based on the Cambridge examinations in English for Foreign Students, being a report circulated to oral examiners and local examiners for those examinations. Cambridge: Local Examinations Syndicate.
- Robins, K. (1996). Interrupting identities: Turkey/Europe. In S. Hall & P. du Gay (Eds.), *Questions of cultural identity* (30-52). London: Sage.
- Rogers, C.R. (1965). The place of the person in the new world of the behavioral sciences. In F.T. Severin (Ed.), *Humanistic viewpoints in psychology: a book of readings* (387-407). New York, NY: McGraw-Hill.
- Ross, S. & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14(2), 159-176.
- Sak, G. (2008). An investigation of the validity and reliability of the speaking exam at a Turkish university. (Unpublished master's thesis). Middle East Technical University, School of Social Sciences, Ankara, Turkey.
- Sarıçoban, G. & Sarıçoban, A. (2012). Atatürk and the history of foreign language education. *The Journal of Language and Linguistic Studies*, 8(1), 24-49.
- Savignon, S.J. (1972). *Communicative competence: an experiment in foreign language teaching*. Philadelphia: The Centre for Curriculum Development, Inc.
- Scriven, M. (1978). How to anchor standards. *Journal of Educational Measurement*, 15, 273-275.
- Seels, B. (1989). The instructional design movement in educational technology. *Educational Technology*, 29(5), 11-15.
- Seidlhofer, B. (2001). Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, 11(2), 133-158.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.

- Sharp, A. (2012). Humanistic approaches to learning. In N.M. Seel (Ed.), *Encyclopedia of the sciences of learning* (1469-1472). Boston: Springer.
- Sheng-ping, T. & Chong-ning, X. (2004, May). On washback of testing to general English education. The Fourth International Conference in ELT, China.
- Shohamy, E. (1988). A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition*, 10, 165-179.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing*, 11, 99-123.
- Shohamy, E., Donitsa-Schmidt, S. & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13, 298-317.
- Sinwongsuwat, K. (2012). Rethinking assessment of Thai EFL learners' speaking skills. *Language Testing in Asia*, 2(4), 75-85.
- Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1-14.
- Skehan, P. and Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task-based learning. *Language Teaching Research*, 1(3), 185-211.
- Sook, K.H. (2003). Korean junior secondary school English teachers' perceptions of speaking assessment. *Asian EFL Journal*, 5(4), 1-30.
- Spada, N. (2007). Communicative language teaching: current status and future prospects. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (271-288). New York, NY: Springer. Doi: [http://dx.doi.org/10.1007/978-0-387-46301-8\\_20](http://dx.doi.org/10.1007/978-0-387-46301-8_20)
- Special Regulation for Anatolian High Schools Which Teach Some Lessons through a Foreign Language, Galatasaray High School - İstanbul Erkek High School [Bir Kısım Derslerin Öğretimini Yabancı Dille Yapan Anadolu Liseleriyle Galatasaray Lisesi – İstanbul Erkek Lisesi Özel Yönetmeliği]. (1976). T.C. Resmi Gazete [T.R. Official Gazette], 15747, 27 October 1976. Retrieved from <http://www.resmigazete.gov.tr/default.aspx>.
- Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing*, 2(1), 31-40.
- Spolsky, B. (1990). Oral examinations: an historical note. *Language Testing*, 7(2), 158-173.
- Spolsky, B. (2008). Language testing at 25: maturity and responsibility. *Language Testing*, 25(3), 297-305.
- Stevenson, D.K. (1985). Authency, validity and a tea party. *Language Testing*, 2(1), 41-47.

- Stiggins, R. & Faires-Conkin, N. (1992). *In teachers' hands*. Albany, NY: State University of New York Press.
- Swain, M. (2001). Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18(3), 275–302. Retrieved from <http://dx.doi.org/10.1177/026553220101800302>.
- Şallı-Çopur, D. (2002). Testing first year FLE students' oral performance using four speaking test methods in spoken English II course and students' attitudes towards these speaking test methods. (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.
- Tarone, E. (1982). Systematicity and attention in interlanguage. *Language Learning*, 32, 9-84.
- Taylor, L. (2001). The paired speaking test format: recent studies. *Cambridge ESOL Research Notes*, 6, 15-17.
- Taylor, L. (2005). Key concepts in ELT: washback and impact. *ELT Journal*, 59(2), 154-155. Retrieved from <http://eltj.oxfordjournals.org/content/59/2/154.full.pdf>.
- Tuan, L.T. (2012). Teaching and assessing speaking performance through analytic scoring approach. *Theory and Practice in Language Studies*, 2(4), 673-679.
- Underhill, A. (1989). Process in humanistic education. *ELT Journal*, 43(4), 250-260.
- Underhill, N. (1987). *Testing spoken language: a handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Ur, P. (1996). *A course in language teaching: practice and theory*. Cambridge: Cambridge University Press.
- Vasuhi, R. (2011, February). Humanism: a human perspective in English language teaching. The International Seminar on Humanistic Language and Literature Teaching, The Anna University, Chennai.
- Vygotsky, L. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.
- Wall, D. (1996). Introducing new tests into traditional systems: insights from general education and from innovation theory. *Language Testing*, 13, 334–354.
- Wang, B. (2010). On rater agreement and rater training. *English Language Teaching*, 3(1), 108-112.
- Wang, G. (2005). Humanistic approach and affective factors in foreign language teaching. *Sino-US English Teaching*, 2(5), 1-5. doi:10.1207/s15430421tip4102\_2
- Weber, R.P. (1990). *Basic content analysis*. Newbury Park, CA: Sage Publications.
- Weir, C.J. (1990). *Communicative language testing*. New Jersey: Prentice Hall.

- Weir, C.J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.
- Weir, C.J. (2003). A Survey of the history of the certificate of proficiency in English (CPE) in the 20th Century. In C.J. Weir & M. Milanovic (Eds.), *Continuity and innovation: revising the Cambridge proficiency in English examination 1913–2002* (1-56). Cambridge: Cambridge University Press.
- Weir, C.J. (2005). *Language testing and validation*. New York: Palgrave Macmillan.
- Widdowson, H.G. (1983). *Learning purpose and language use*. Oxford: Oxford University Press.
- Widdowson, H.G. (1994). The ownership of English. *TESOL Quarterly*, 28, 377-389.
- Wiggins, G. (1998). *Educative assessment: designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14(1), 85-106.
- Wigglesworth, G. and Frost, K. (2017). Task and performance-based assessment. In E. Shohamy and N.H. Hornberger (Eds.), *Encyclopedia of language and education: language testing and assessment* (Third Edition, 121-134). Cham, Switzerland: Springer International Publishing.
- Wilkinson, A. (1968). The testing of oracy. In A. Davies (Ed.), *Language testing symposium* (117-132). Oxford: Oxford University Press.
- Willis, D. & Willis, J. (2007). *Doing task-based teaching*. Oxford, England: Oxford University Press.
- Wilson, B.G. (Ed.). (1996). *Constructivist learning environments: case studies in instructional design*. Englewood Cliffs: Educational Technology Pub. Inc.
- Xiaotong, H. (2014). An analysis of communicative language teaching methodology and its application on language learning in China. *Higher Education of Social Science*, 6(3), 100-105. doi: 10.3968/4709
- Young, S. & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14(4), 403-424.
- Zhang, Y. & Wildemuth, B.M. (2005). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (308-319). Westport, CT: Libraries Unlimited.

**APPENDICES**

### APPENDIX A: Sample of Holistic Scale

<b>Proficient User</b>	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
<b>Independent User</b>	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
<b>Basic User</b>	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

*Note:* It is retrieved from Council of Europe, 2001, p. 24.



### APPENDIX B: Sample of Analytic Scale

Level	Range	Accuracy	Fluency	Interaction	Coherence	Phonology
<b>C2</b>	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it.	Can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly. Can interweave his/her contribution into the joint discourse with fully natural turntaking, referencing, allusion making etc.	Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.	Can employ the full range of phonological features in the target language with a high level of control – including prosodic features such as word and sentence stress, rhythm and intonation – so that the finer points of his/her message are clear and precise. Intelligibility is not affected in any way by features of accent that may be retained from other language(s).
<b>C1</b>	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say.	Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can select a suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/her own contributions skilfully to those of other speakers.	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.	Can employ the full range of phonological features in the TL with sufficient control to ensure intelligibility throughout. Can articulate virtually all the sounds of the TL; some features of accent retained from other language(s) may be noticeable, but they do not affect intelligibility at all.

Level	Range	Accuracy	Fluency	Interaction	Coherence	Phonology
<b>B2</b>	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses.	Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting others in, etc.	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some "jumpiness" in a long contribution.	Can generally use appropriate intonation, place stress correctly and articulate individual sounds clearly; accent tends to be influenced by other language(s) he/she speaks, but has little or no effect on intelligibility.
<b>B1</b>	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.	Pronunciation is generally intelligible; can approximate intonation and stress at both utterance and word levels. However, accent is usually influenced by other language(s) he/she speaks.

Level	Range	Accuracy	Fluency	Interaction	Coherence	Phonology
<b>A2</b>	Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Uses some simple structures correctly, but still systematically makes basic mistakes.	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulations are very evident.	Can ask and answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can link groups of words with simple connectors like "and", "but" and "because".	Pronunciation is generally clear enough to be understood, but conversational partners will need to ask for repetition from time to time. A strong influence from other language(s) he/she speaks on stress, rhythm and intonation may affect intelligibility, requiring collaboration from interlocutors. Nevertheless, pronunciation of familiar words is clear.
<b>A1</b>	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorised repertoire.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.	Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	Can link words or groups of words with very basic linear connectors like "and" or "then".	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by interlocutors used to dealing with speakers of the language group concerned. Can reproduce correctly a limited range of sounds as well as the stress on simple, familiar words and phrases.

*Note:* It is retrieved from Council of Europe, 2018, pp. 171-172.

## APPENDIX C: Teacher Questionnaire

### TEACHER QUESTIONNAIRE

Dear Colleague,

This questionnaire has been designed to get attitudes and perceptions of ELT-teachers on speaking tests, which is a part of my thesis, namely, an investigation of assessing speaking skill for the achievement tests of the 9<sup>th</sup> grade students in Anatolian high schools. Your sincere responses are very important for validity and reliability of the study. The responses you give will be confidential and analyzed by taking your privacy into consideration. After you give your demographic information, please fill in the Part 1 and Part 2. Please do not skip any items.

Thank you for your contribution.

Ceyda ÖZDEMİR  
English Teacher

Assoc. Prof. Dr. Turan PAKER  
Pamukkale University

#### DEMOGRAPHIC INFORMATION

**1. Gender:** .....

**2. Age:** .....

**3. BA program you have graduated:**

- a) English Language Teaching
- b) English Language and Literature / American Culture and Literature
- c) Translation and Interpretation
- d) Comparative Literature
- e) Linguistics
- f) Other (.....)

**4. MA degree:**    a) Yes    b) No

- a) ELT.
- b) Educational Sciences.
- c) English Language and Literature
- d) Other (.....)
- e) I have been doing MA in .....

**5. Experience in teaching:** ..... years

**PART 1**

Put a cross (x) to the choice which represents your idea/attitude by reading items below.

	(1)Strongly disagree	(2)Disagree	(3)Partly agree	(4)Agree	(5)Strongly agree
1. I have prepared a speaking test before.					
2. I have administered a speaking test before.					
3. I am inexperienced with speaking tests.					
4. I plan the process of the speaking test with my colleagues at the beginning of the term or academic year.					
5. I prepare a rubric or a checklist for the speaking test.					
6. My colleagues and I agreed on rubric items together.					
7. I let the students know what testing procedure will be adopted prior to a speaking test.					
8. I declare rubric items and marks for each item before a speaking test.					
9. I give extra materials to the students before the speaking test so that they can study better.					
10. I give the necessary time for the students in order to cram for the speaking test by announcing the time of it in advance.					
11. I have some worries about the administration of a speaking test before.					
12. I tested the students in a room which is quiet.					
13. I tested the students in the classroom with the other students.					
14. The classroom was noisy during the speaking test.					
15. It is better to assign only one <b>interlocutor*</b> and one <b>assessor*</b> as graders.					
16. I just listened to the students and gave their marks as an assessor during the speaking test.					
17. I both asked questions to the students and gave their marks as both assessor and interlocutor during the speaking test.					
18. I combined speaking skill with another skill (listening, reading or writing).					
19. The time was enough for each student.					
20. I feel more anxious and nervous during the speaking test in contrast with pen-and-paper tests.					
21. The students' anxiety level is very high during the speaking test.					
22. I tolerated their mistakes not to demotivate the students during the speaking test.					
23. I supported the students with my positive behavior during the speaking test.					
24. I behaved in a biased way for some students during the speaking test.					
25. I used a checklist or a rubric to assess their performance.					
26. I recorded the students' performances.					
27. The tasks and activities were not above the students' abilities and levels.					
28. The tasks and activities were parallel with those in the course book.					
29. The tasks and activities were clear and comprehensible.					
30. The tasks and activities were from daily life.					
31. Some students whose English level is high cannot perform well due to their test anxiety during the speaking test.					
32. I gave each student's mark during the speaking test.					
33. I gave each student's mark at the end of the speaking test after all students had finished it.					
34. Rubric items for the speaking test were clear.					

\***Interlocutor**: a person who is having a conversation with students.

\***Assessor**: a person who is listening to students and giving marks for their performance.

	(1) Strongly disagree	(2) Disagree	(3) Partly agree	(4) Agree	(5) Strongly agree
35. The rubric for the speaking test was satisfactory to assess them effectively.					
36. I was objective in assessing the performance of the students.					
37. I think I am a reliable rater.					
38. I know that the speaking test is compulsory at high schools.					
39. Speaking skills can be assessed accurately.					
40. The speaking test is necessary to see the students' speaking level.					
41. The speaking test encourages students to use their speaking ability.					
42. The speaking test helps the students to notice the weaknesses in their speaking performances.					
43. Trying to test speaking skill is a waste of time.					
44. If speaking skills were not tested, I would not spend so much time on teaching the speaking skill.					
45. It is difficult for me to prepare the speaking test.					
46. It is stressful to administer the speaking test for me.					
47. All teachers should use the rubric developed and imposed by the Ministry of National Education.					
48. It is difficult to prepare rubric for the speaking test.					
49. Rubric is useless in the speaking test.					
50. It is difficult to be objective as marking the speaking test for me.					
51. I believe that all the teachers at my school are objective.					
52. All teachers should be trained through in-service training on how speaking skill can be tested.					

**PART 2**

**Circle a choice which represents your idea/attitude and fill in the blanks if it is necessary.**

1) Have you ever taken part in interlocutor training sessions before speaking tests? Where?

- a) Yes (write here, please) ..... b) No

2) Did you plan timing before a speaking test? How much time did you allocate for each student during a speaking test?

- a) Yes (write here, please) ..... b) No

3) Which one or ones of them affected you while preparing a speaking test? You can circle more than one choice.

- |                       |                        |
|-----------------------|------------------------|
| a) Students' level    | e) English course book |
| b) The curriculum     | f) All                 |
| c) My own wish        | g) Other (.....)       |
| d) Students' interest |                        |

4) How did you support your students to improve their speaking skills and practice or get prepared for speaking tests? You can circle more than one choice.

- a) I suggest extra sources such as video, film, etc.
- b) I do speaking activities during the class.
- c) I encourage them to speak.
- d) I motivate the students to speak English except class hours.
- e) All
- f) None
- g) Other (.....)

5) How many times do the students take speaking test in a term?

- a) 1
- b) 2
- c) 3
- d) more than 3
- e) none

6) How many teachers test speaking skill of the students?

- a) 1
- b) 2
- c) 3
- d) more than 3

7) When did you test the students? You can circle more than one choice.

- a) During the class time
- b) In my free time
- c) After school time
- d) During English lessons
- e) In your student's free time
- f) Other (.....)

8) How did you test your students? You can circle more than one choice.

- a) Individually
- b) In pairs
- c) In groups
- d) All of them
- e) Other (.....)

9) What kind of tests do you think are more advantageous to use while testing speaking skills? You can circle more than one choice.

- a) Reading aloud
- b) Sentence repetition
- c) Sentence transformation
- d) Sentence construction
- e) Sentence completion
- f) Completing a dialogue
- g) Translating/Interpreting a text/dialogue
- h) Information-gap activities
- i) Talking about pictures
- j) Talking about video / film
- k) Question and answer
- l) Giving instructions / explanation / description
- m) Problem solving activity
- n) Role-playing
- o) Verbal essay
- p) Oral interview
- q) Discussion
- r) Oral presentation
- s) Students write a short story / paragraph or complete a dialogue
- t) All
- u) Other (.....)

Why?

.....

.....

.....

10) What kind of materials did you use during the speaking test? You can circle more than one choice.

- a) Course book
- b) Pictures
- c) Objects, realia
- d) Films, video
- e) EBA software
- f) Question pool
- g) Topics for discussion
- h) News, magazines, brochures
- i) Dialogue for role play
- j) Incomplete stories/reading passages / dialogues
- k) Topics for presentation

11) What type of a rating scale did you prefer while testing students' speaking skills?

- a) **Holistic (impressionistic)\***
- b) **Analytic\***
- c) Both
- d) None

Why?

.....

.....

.....

\***Holistic (impressionistic) Scoring:** you score to performance on the basis of an overall impression of it.

\***Analytic Scoring:** you score each of aspects of a task separately and generally use a scale such as rubric, checklist, etc.

12) What did you take into consideration related to the following criteria below while testing speaking? You can circle more than one choice.

- a) Clarity of expression
- b) Content
- c) Organization
- d) Vocabulary
- e) Accuracy
- f) Fluency
- g) Pronunciation
- h) Body language
- i) Eye contact
- j) All
- k) None
- l) Your own criteria (.....)



13) What are great challenges in the assessment of speaking skills in your opinion? You can circle more than one choice.

- a) It is time-consuming.
- b) Physical factors (Classroom, organization, time, noise) affect the test success.
- c) It is difficult to prepare a rating scale.
- d) Speaking tests have validity and reliability problems.
- e) It needs expertise.
- f) Speaking tests do not yield objective scores.
- g) Listening to and dealing with so many students is tiring and weary.
- h) Giving clue or asking extra questions in order to prompt students to speak affects objectivity.
- i) There is no proper standard speaking test model.
- j) It needs experience.
- k) All
- l) None
- m) Other (.....)

Why?

.....  
.....

14) How should speaking skills be tested?

- a) By continuous assessment (all the time in class).
- b) By achievement tests every month.
- c) By an achievement test at the end of each term.
- d) By using all the suggestions mentioned above.
- e) Other (.....)

## APPENDIX D: Students Questionnaire

### STUDENTS QUESTIONNAIRE

Dear student,

This questionnaire has been designed to get attitudes and perceptions of the 9<sup>th</sup> grade students on speaking tests, which is a part of my thesis, namely, an investigation of assessing speaking skill for the achievement tests of the 9<sup>th</sup> grade students in Anatolian high schools. Your sincere responses are very important for validity and reliability of the study. The responses you give will be confidential and analyzed by taking your privacy into consideration. After you give your demographic information, please fill in the Part 1 and Part 2. Please do not skip any items.

Thank you for your contribution.

Ceyda ÖZDEMİR  
English Teacher

Assoc. Prof. Dr. Turan PAKER  
Pamukkale University

#### DEMOGRAPHIC INFORMATION

Age: .....

Gender: .....

#### PART 1

**Put a cross (x) to the choice which represents your idea/attitude by reading items below.**

	(1)Strongly disagree	(2)Disagree	(3)Partly agree	(4)Agree	(5)Strongly agree
1. I was informed about testing procedure prior to the speaking test.					
2. I was informed about how my performance would be assessed prior to the speaking test.					
3. My teachers gave me extra materials to be able to study better before the speaking test.					
4. I had sufficient time to study for the speaking test.					
5. I was inexperienced with how I should get ready for a speaking test.					
6. I had no idea about the grading of the speaking test.					
7. I took speaking test alone in the classroom.					
8. I took speaking test with my one or two friends as pair or group in the classroom.					
9. As I was taking speaking test, the other students were also in the classroom.					
10. It discomforted me in terms of attention, tension, noise, etc. because the other students were also in the classroom as I was taking speaking test.					
11. My teachers' attitude toward me influenced my performance negatively during the speaking test.					

	(1)Strongly disagree	(2)Disagree	(3)Partly agree	(4)Agree	(5)Strongly agree
12. My teachers gave me sufficient time to think about my performance during the speaking test.					
13. I had difficulties in understanding pronunciation of my teacher.					
14. It is hard to express myself clearly during the speaking test.					
15. My personal characteristics such as coyness, timidity, nervousness, etc. stop me from speaking English during the test.					
16. My teachers encouraged and supported me to perform better during the speaking test.					
17. My teachers gave me feedback related to my performance in the speaking test.					
18. The speaking test was recorded.					
19. The tasks and activities were loud and clear in the speaking test.					
20. The speaking test was parallel with class activities.					
21. The tasks and activities were from daily life in the speaking test.					
22. Some tasks and activities were non-class and unexpected in the speaking test.					
23. The tasks and activities were difficult.					
24. Speaking activities during lessons were not sufficient.					
25. I wanted to do more speaking activities in class after this test.					
26. The time was not sufficient for each student in speaking test.					
27. I notice my weaknesses in speaking after speaking test.					
28. I feel that I need to study English harder after speaking test.					
29. I performed better than I expected in speaking test.					
30. I notice my strengths in speaking after speaking test.					
31. My teachers were objective at scoring.					
32. My teachers assessed my performance on the basis of a rubric or a checklist.					
33. The grade I got from speaking test correctly reflected my speaking ability.					
34. Speaking test is essential to identify my speaking level.					
35. I feel more relaxed in speaking tests in contrast with pen-and-paper tests.					
36. If I did not have to take a speaking test, I would not take part in speaking activities in the lessons.					
37. Speaking test is useful to improve my English.					
38. I can also use many of the things in lessons which I have studied for the speaking test.					
39. If I did not have to take a speaking test, I would not spend so much time to improve my speaking skill.					
40. I think I need to take speaking tests more frequently.					
41. Speaking test makes me nervous.					
42. Speaking test is the most difficult test.					
43. Speaking test is a test that I find myself the most unsuccessful.					
44. It is waste of time to assess speaking skill.					

## PART 2

**Circle a choice which represents your idea/attitude and fill in the blanks if it is necessary.**

1) Have you ever taken a speaking test before? Where?

a) Yes (write here, please).....

b) No

2) What did you do to get ready for the speaking test? You can circle more than one choice.

- |  |   |
|--|---|
| a) I followed the lesson carefully.                              | e) I practiced conversation in my daily life. |
| b) I crammed for the test on my own.                             | f) All  |
| c) I crammed for the test with my friends.                       | g) None                                       |
| d) I got professional help (language course or English teacher). | h) Other (.....)                              |

3) How many teachers were there in the speaking test?

- a) 1      b) 2      c) 3      d) more than 3

4) When do you take your speaking test at your school? You can circle more than one choice.

- |                          |                              |
|--------------------------|------------------------------|
| a) During the class time | d) During English lessons    |
| b) In my free time       | e) In my teacher's free time |
| c) After school time     | f) Other (.....)             |

5) Did your teachers use any material during the speaking test?

- a) Yes (write here, please) .....      b) No

6) Which activities did you do in this speaking test? You can circle more than one choice.

- |   |   |
|---|---|
| a) Making a presentation on a topic                   | f) Acting out dialogues with another classmate                  |
| b) Talking about pictures                             | g) Discussing in groups   |
| c) Introducing yourself and your family               | h) Role-playing   |
| d) Answering questioning randomly from question pouch | i) Talking about video / film (character, pilot, setting, etc.) |
| e) Answering questions given before the test          | j) Talking about something he listens or reads                  |
|   | k) Other (.....)  |

7) Which difficulties did you have in speaking test? You can circle more than one choice.

- |   |                          |
|---|--------------------------|
| a) Understanding questions / activities                               | f) Pronouncing correctly |
| b) Answering appropriately to questions                               | g) Contextualizing       |
| c) First thinking in Turkish, then translating to English as speaking | h) Speaking accurately   |
| d) Speaking fluently  | i) Overcoming the stress |
| e) Finding correct vocabulary as speaking                             | j) All                   |
|   | k) None                  |
|   | l) Other (.....)         |

8) In your opinion, which criteria should be taken into consideration to assess a student in a speaking test? You can circle more than one choice.

- |   |                    |
|---|--------------------|
| a) Content (speak related to content)               | g) Grammar         |
| b) Organization of speech (organization of content) | h) Body language   |
| c) Fluency  | i) Eye contact     |
| d) Clarity of expression                            | j) Self-confidence |
| e) Vocabulary knowledge                             | k) All             |
| f) Correct pronunciation                            | l) None            |
|   | m) Other (.....)   |

## **APPENDIX E: Teacher Interview Questions**

### **INTERVIEW QUESTIONS**

- 1) What aspects of the speaking test do you like most?
- 2) What aspects of the speaking test do you like least?
- 3) Do you have any suggestions on how to improve the procedure of ‘assessment of speaking’?
- 4) Can you briefly describe a good quality speaking test procedure in your opinion?
- 5) What kind of measures do you take to ensure a high level of reliability?
  - a. Do you consider the issues of inter and intra-rater reliability when you test speaking?
  - b. If you give different marks to the same student how do you handle it?

**APPENDIX F: Student Interview Questions****INTERVIEW QUESTIONS**

- 1) What aspects of the speaking test do you like most?
- 2) What aspects of the speaking test do you like least?
- 3) Are you happy with your grade for the speaking test?
  - a. If you are not, do you talk about it with your teacher?
  - b. How does your teacher deal with it?
- 4) What would you like to change related to speaking test?
- 5) What can else be done in speaking test in your opinion? What do you suggest for it?

**APPENDIX G: Teacher Questionnaire (Turkish Version)**

**ÖĞRETMEN ANKETİ**

Sevgili Meslektaşım,

Anadolu Liseleri'nde 9. Sınıf öğrencilerine uygulanan konuşma sınavlarının değerlendirilmesi amacıyla bir araştırma yürütülmektedir. Bu anket, öğretmenlerin konuşma sınavlarıyla ilgili duygu ve düşüncelerinin tespiti için yüksek lisans tezimin bir parçası olarak hazırlanmıştır. Çalışmanın geçerliliği ve güvenilirliği bakımından soruları içtenlikle cevaplamanız çok önemlidir. Cevaplarınız gizli tutulacaktır. Lütfen kişisel bilgileri doldurduktan sonra Bölüm 1 ve Bölüm 2'yi cevaplandırınız. Hiçbir soruyu boş bırakmamanızı rica ederiz.

Katkılarınız için teşekkür ederiz.

Ceyda ÖZDEMİR  
İngilizce Öğretmeni

Doç. Dr. Turan PAKER  
Pamukkale Üniversitesi

---

**KİŞİSEL BİLGİLER**

**1. Cinsiyet:** .....

**2. Yaş:** .....

**3. Mezun olduğunuz lisans programı:**

- a) İngiliz Dili Eğitimi
- b) İngiliz Dili ve Edebiyatı / Amerikan Kültür ve Edebiyatı
- c) Mütercim Tercümanlık
- d) Karşılaştırmalı Edebiyat
- e) Dilbilim
- f) Diğer (.....)

**4. Yüksek lisans Programı:**      a) Evet                      b) Hayır

- a) İngiliz Dili Eğitimi
- b) Eğitim Bilimleri
- c) İngiliz Dili ve Edebiyatı
- d) Diğer (.....)
- e) Hala ..... programında yüksek lisans öğrencisiyim

**5. Öğretmenlik deneyimi:** ..... yıl



**BÖLÜM 1**

Aşağıdaki maddeleri dikkatlice okuyarak size uygun gelen seçeneğe (x) işareti koyunuz.

	(1) Kesinlikle katılmıyorum	(2) Katılmıyorum	(3) Kısmen Katılıyorum	(4) Katılıyorum	(5) Kesinlikle katılıyorum
1. Daha önce konuşma sınavı hazırladım.					
2. Daha önce konuşma sınavı uyguladım.					
3. Konuşma sınavlarında yeterince deneyimim yok.					
4. Dönem ya da sene başında meslektaşlarımla konuşma sınavını planlarım.					
5. Konuşma sınavı için dereceli puanlama anahtarı ya da kontrol listesi hazırlarım.					
6. Ben ve zümrem değerlendirmede kullanılacak dereceli puanlama anahtarının ölçütlerini beraber kararlaştırırız.					
7. Konuşma sınavı öncesinde sınav süreciyle ilgili olarak öğrencileri bilgilendiririm.					
8. Konuşma sınavı öncesinde öğrenciye, değerlendirmede kullanılacak dereceli puanlama anahtarını tanıtır ve beklentilerimi yansıtırım.					
9. Konuşma sınavı öncesi sınava daha iyi hazırlanabilmeleri için öğrencilere ek materyal veririm.					
10. Konuşma sınavının zamanını önceden bildirerek öğrencinin sınava hazırlanması için gerekli süreyi veririm.					
11. Öncesinde konuşma sınavının uygulanışı hakkında endişelerim vardı.					
12. Öğrencileri sessiz bir odada sınava aldım.					
13. Öğrencileri sınıfta diğer öğrencilerle birlikte sınav yaptım.					
14. Sınav yeri gürültülüydü.					
15. Sınavda bir soru soran ve bir de değerlendirme yapan öğretmen olması daha etkilidir.					
16. Sınavda sadece değerlendirme yapan öğretmen olarak öğrenciyi dinledim ve öğrencinin notunu verdim.					
17. Sınavda hem öğrenciye soru sordum hem de öğrencinin notunu verdim.					
18. Konuşma becerisini diğer becerilerle (dinleme, okuma ve yazma) birleştirerek sordum.					
19. Her öğrenci için süre yeterliydi.					
20. Konuşma sınavında, yazılı sınavlara kıyasla kendimi daha gergin hissediyorum.					
21. Sınavda öğrencilerin stres seviyesi çok yüksekti.					
22. Öğrencilerin şevkini kırmamak için hatalarını görmezden geldim.					
23. Sınavda pozitif tutumumla öğrencileri destekledim.					
24. Sınavda bazı öğrencilere önyargılı davrandım.					
25. Öğrencileri değerlendirmek için dereceli puanlama anahtarı ya da kontrol listesi kullandım.					
26. Konuşma sınavını video ya da ses kaydıyla kaydettim.					
27. Etkinlikler öğrencilerin seviyesinin üstünde değildi.					
28. Sınavdaki etkinlikler kitaptaki etkinliklerle paraleldi.					
29. Etkinlikler net ve anlaşılırdı.					
30. Etkinlikler günlük hayattandı.					
31. Sınavda İngilizce seviyesi yüksek bazı öğrenciler, sınav stresi sebebiyle iyi performans sergileyemediler.					
32. Her öğrencinin notunu konuşma sınavı esnasında verdim.					

	(1) Kesinlikle katılmıyorum	(2) Katılmıyorum	(3) Kısmen Katılıyorum	(4) Katılıyorum	(5) Kesinlikle katılıyorum
33. Bütün öğrencilerin notlarını sınav bittikten sonra verdim.					
34. Dereceli puanlama anahtarının ölçütleri netti.					
35. Dereceli puanlama anahtarı, öğrencileri etkili bir şekilde değerlendirmek için tatmin edici bir araçtı.					
36. Öğrenci performanslarını değerlendirirken objektiftim.					
37. Ben tutarlı ve güvenilir bir değerlendirme yaptığıma inanıyorum.					
38. Konuşma sınavının ortaöğretim kurumlarında zorunlu olduğunu biliyorum.					
39. Konuşma becerisi kesin olarak değerlendirilebilir.					
40. Konuşma sınavı, öğrencinin seviyesini görmek için gereklidir.					
41. Konuşma sınavı, öğrenciyi konuşmaya teşvik eder.					
42. Konuşma sınavı, öğrencinin performansındaki eksiklerini fark etmesine yardım eder.					
43. Konuşma becerisini değerlendirmeye çalışmak zaman kaybı.					
44. Konuşma becerisi değerlendirilmeseydi, derslerde bu beceri üstünde fazla zaman harcamazdım.					
45. Benim için konuşma sınavı hazırlamak zordur.					
46. Konuşma sınavını uygulamak benim için streslidir.					
47. Bütün öğretmenler, Milli Eğitim Bakanlığı'nın hazırladığı standart bir dereceli puanlama anahtarı kullanmalıdır.					
48. Dereceli puanlama anahtarı hazırlamak zordur.					
49. Konuşma sınavında dereceli puanlama anahtarı gereksizdir.					
50. Benim için konuşma sınavını puanlarken objektif olmak zordur.					
51. Okuldaki tüm öğretmenlerin objektif olduğuna inanıyorum.					
52. Bütün öğretmenler, konuşma becerisinin nasıl değerlendirilmesi gerektiği üzerine hizmet içi eğitim almalıdır.					

## BÖLÜM 2

**Aşağıdaki soruları dikkatlice okuyarak size uygun seçeneği işaretleyiniz ve boşlukları istenen bilgi ile doldurunuz.**

1) Konuşma sınavından önce konuşma sınavının değerlendirilmesiyle ilgili eğitim aldınız mı?

a) Evet (ise yerini yazın) .....

b) Hayır

2) Konuşma sınavından önce her öğrenciye ayrılacak süreyi planlar mısınız? Sınav esnasında her öğrenciye ne kadar süre ayırırsınız?

a) Evet (ise süreyi yazın) .....

b) Hayır

3) Konuşma sınavı hazırlarken sizi hangi kriterler etkiler? Birden fazla seçeneği işaretleyebilirsiniz.

- |                          |                          |
|--------------------------|--------------------------|
| a) Öğrencilerin seviyesi | e) İngilizce ders kitabı |
| b) Müfredat              | f) Hepsi                 |
| c) Kendi isteğim         | g) Diğer (.....)         |
| d) Öğrencilerin ilgileri |                          |

4) Öğrencilerin konuşma becerilerini geliştirmek, pratik yapmalarını sağlamak ya da konuşma sınavına hazırlanmalarına yardım etmek için neler yaparsınız? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |  |
|---|--|
| a) Video, film gibi ek kaynaklar öneririm.  | d) Öğrencileri, sınıf dışında İngilizce konuşmaya teşvik ederim. |
| b) Sınıfta konuşma etkinlikleri yaparım.    | e) Hepsi   |
| c) Öğrencileri konuşmaya cesaretlendiririm. | f) Hiçbiri   |
|   | g) Diğer (.....)   |

5) Öğrencilerinize bir dönemde kaç kez konuşma sınavı uygularsınız?

- |      |      |      |                |        |
|------|------|------|----------------|--------|
| a) 1 | b) 2 | c) 3 | d) 3'ten fazla | e) Hiç |
|------|------|------|----------------|--------|

6) Öğrencinin konuşma becerisini kaç tane öğretmen değerlendirir?

- |      |      |      |                |
|------|------|------|----------------|
| a) 1 | b) 2 | c) 3 | d) 3'ten fazla |
|------|------|------|----------------|

7) Öğrencileri ne zaman sınav yaparsınız? Birden fazla seçeneği işaretleyebilirsiniz.

- |                                |                              |
|--------------------------------|------------------------------|
| a) Herhangi bir ders esnasında | d) İngilizce dersi esnasında |
| b) Boş zamanımda               | e) Öğrencinin boş zamanında  |
| c) Okuldan sonra               | f) Diğer (.....)             |

8) Öğrencilerinizi nasıl değerlendirirsiniz? Birden fazla seçeneği işaretleyebilirsiniz.

- |                         |                  |
|-------------------------|------------------|
| a) Bireysel             | d) Hepsinde      |
| b) İkili etkinliklerde  | e) Diğer (.....) |
| c) Grup etkinliklerinde |                  |

9) Konuşma becerisini değerlendirirken sınavda hangi etkinlikler daha avantajlıdır? Birden fazla seçeneği işaretleyebilirsiniz.

- |                                  |                                       |
|----------------------------------|---------------------------------------|
| a) Yüksek sesle okumak           | l) Bilgi verme / açıklama / betimleme |
| b) Cümleyi tekrar etme           | m) Problem çözme etkinlikleri         |
| c) Cümle yapısı değiştirme       | n) Rol yapma                          |
| d) Cümle kurma                   | o) Bir konu üzerine tartışma          |
| e) Cümle tamamlama               | p) Röportaj yapma                     |
| f) Diyalog tamamlama             | q) Tartışma grubu                     |
| g) Bir metni / diyalogu çevirme  | r) Sunum                              |
| h) Boşluk doldurma etkinlikleri  | s) Kısa hikaye / paragraf yazma       |
| i) Resim hakkında konuşma        | t) Hepsi                              |
| j) Video / film hakkında konuşma | u) Diğer (.....)                      |
| k) Soru - Cevap                  |                                       |

Neden?

.....  
 .....

10) Konuşma sınavında hangi materyalleri kullanırsınız? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |   |
|---|---|
| a) Ders kitabı                                      | g) Tartışma konuları                              |
| b) Resimler   | h) Gazete, dergi, broşür                          |
| c) Nesnelere, gerçek nesnelere (eğitsel malzemeler) | i) Canlandırma için diyaloglar                    |
| d) Film, video                                      | j) Tamamlanmamış hikaye / okuma parçası / diyalog |
| e) EBA yazılımı                                     | k) Sunum için konular                             |
| f) Soru havuzu                                      |   |

11) Öğrencilerin konuşma becerilerini değerlendirirken ne çeşit bir ölçek kullanmayı tercih edersiniz?

- |                                   |             |             |            |
|-----------------------------------|-------------|-------------|------------|
| a) Bütünsel (İzlenimsel-Holistik) | b) Analitik | c) İkisi de | d) Hiçbiri |
|-----------------------------------|-------------|-------------|------------|

Neden?

.....  
 .....

12) Konuşma sınavında aşağıdaki kriterlerden hangilerini göz önünde bulundurursunuz? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |                             |
|---|-----------------------------|
| a) Anlamsal netlik                                  | g) Doğru telaffuz           |
| b) İçerik   | h) Vücut dili               |
| c) Konuşmanın organizasyonu (içeriğin düzenlenmesi) | i) Göz teması               |
| d) Kelime bilgisi                                   | j) Hepsi                    |
| e) Dil bilgisi                                      | k) Hiçbiri                  |
| f) Konuşmanın akıcılığı                             | l) Sizin kriteriniz (.....) |

**13)** Konuşma becerilerini değerlendirmedeki sorunlar sizce nelerdir? Birden fazla seçeneği işaretleyebilirsiniz.

- a) Çok fazla zaman alır.
- b) Fiziksel etmenler (Sınıf, organizasyon, zamanı ayarlama, gürültü) sınav başarısını etkiler.
- c) Değerlendirme ölçeği hazırlamak zordur.
- d) Konuşma sınavlarında geçerlilik ve güvenilirlik problemleri vardır.
- e) Uzmanlık ister.
- f) Konuşma sınavları nesnel puanlar vermez.
- g) Bir oturumda çok fazla öğrenciyi sınav yapmak yorucudur.
- h) Bazı öğrencileri konuşmaya teşvik için ipucu verme veya ek sorular sorma tarafsızlığı etkiler.
- i) Standart bir konuşma sınavı modeli yoktur.
- j) Deneyim ister.
- k) Hepsi
- l) Hiçbiri
- m) Diğer (.....)

Neden?

.....  
 .....

**14)** Sizce konuşma becerileri nasıl değerlendirilmelidir?

- a) Sınıf içinde devamlı değerlendirilmeli
- b) Her ay yapılan sınavlarla değerlendirilmeli
- c) Dönem sonundaki bir sınavla değerlendirilmeli
- d) Yukarıdaki önerilerin hepsi
- e) Diğer

## APPENDIX H: Students Questionnaire (Turkish Version)

### ÖĞRENCİ ANKETİ

Sevgili Öğrenci,

Anadolu Liselerinde 9. sınıf öğrencilerine uygulanan konuşma sınavlarının değerlendirilmesi amacıyla bir araştırma yürütülmektedir. Bu anket, öğrencilerin konuşma sınavlarıyla ilgili duygu ve düşüncelerinin tespiti için yüksek lisans tezimin bir parçası olarak hazırlanmıştır. Çalışmanın geçerliliği ve güvenilirliği bakımından soruları içtenlikle cevaplamanız çok önemlidir. Lütfen kişisel bilgileri doldurduktan sonra Bölüm 1 ve Bölüm 2'yi cevaplandırınız. Hiçbir soruyu boş bırakmamanızı rica ederiz.

Katkılarınız için teşekkür ederiz.

Ceyda ÖZDEMİR  
İngilizce Öğretmeni

Doç. Dr. Turan PAKER  
Pamukkale Üniversitesi

### KİŞİSEL BİLGİLER

Yaş: .....

Cinsiyet: .....

### BÖLÜM 1

Aşağıdaki maddeleri dikkatlice okuyarak size uygun gelen seçeneğe (x) işareti koyunuz.

	(1) Kesinlikle katılmıyorum	(2) Katılmıyorum	(3) Kısmen Katılıyorum	(4) Katılıyorum	(5) Kesinlikle katılıyorum
1. Konuşma sınavının hangi bölümlerden oluşacağı hakkında bilgilendirilmiştim.					
2. Konuşma sınavından önce değerlendirmenin nasıl yapılacağı hakkında bilgilendirilmiştim.					
3. Konuşma sınavından önce öğretmenlerim daha iyi çalışabilmem için bana fazladan materyal verdi.					
4. Konuşma sınavına çalışmak için yeterince zamanım vardı.					
5. Konuşma sınavına nasıl hazırlanmam konusunda deneyimim yoktu.					
6. Konuşma sınavının puanlaması hakkında en ufak bir fikrim yoktu.					
7. Konuşma sınavında sınıfta tek başıyım.					
8. Konuşma sınavında sınıfta bir-iki arkadaşımın birlikte grup olarak sınav oldum.					
9. Konuşma sınavında, ben (tek) sınav olurken sınıf arkadaşlarım da sınıftaydı.					
10. Konuşma sınavında, ben sınav olurken sınıf arkadaşlarımda da sınıfta olması beni dikkat toplama, heyecan, gürültü vb. bakımından rahatsız etti.					
11. Konuşma sınavı esnasında öğretmenlerimin bana karşı tutumu performansımı olumsuz etkiledi.					

	(1) Kesimlikle katılmıyorum	(2) Katılmıyorum	(3) Kısmen Katılıyorum	(4) Katılıyorum	(5) Kesimlikle katılıyorum
12. Konuşma sınavı esnasında, öğretmenlerim performansım hakkında düşünmem için bana yeterli zamanı verdi.					
13. Konuşma sınavında öğretmenin telaffuzunu anlamakta zorlandım.					
14. Konuşma sınavında kendimi açık bir şekilde ifade etmekte zorlandım.					
15. Konuşma sınavında utangaçlık, çekingenlik, aşırı heyecan gibi kişisel özelliklerim İngilizce konuşmamı engelledi.					
16. Konuşma sınavında öğretmenlerim daha iyisini yapmam için beni cesaretlendirip desteklediler.					
17. Konuşma sınavında öğretmenler performansıma ilişkin bana geri bildirim verdiler.					
18. Konuşma sınavı video ya da ses kaydıyla kaydedildi.					
19. Konuşma sınavındaki etkinlikler açık ve netti.					
20. Konuşma sınavı derste yaptığımız konuşma etkinlikleriyle paraleldi.					
21. Konuşma sınavındaki etkinlikler günlük hayattandı.					
22. Konuşma sınavındaki bazı etkinlikler daha önce hiç karşılaşmadığım ve beklemediğim etkinliklerdi.					
23. Konuşma sınavındaki etkinlikler zordu.					
24. Derste yapılan konuşma etkinlikleri yetersizdi.					
25. Konuşma sınavından sonra sınıfta daha çok konuşma etkinlikleri yapılmasını istedim.					
26. Konuşma sınavının süresi yetersizdi.					
27. Konuşma sınavından sonra konuşmadaki zayıf yönlerimi fark ettim.					
28. Konuşma sınavından sonra İngilizce dersine daha çok çalışmam gerektiğini hissettim.					
29. Konuşma sınavında beklediğimden daha iyi performans sergiledim.					
30. Konuşma sınavından sonra konuşmadaki güçlü yönlerimi fark ettim.					
31. Sınav puanlamasında öğretmenlerim objektifti.					
32. Öğretmenlerim benim performansımı bir ölçek yardımıyla değerlendirdi.					
33. Konuşma sınavından aldığım not benim konuşma becerimi doğru bir şekilde yansıttı.					
34. Konuşma sınavı İngilizce düzeyimi belirlemede gereklidir.					
35. Konuşma sınavındayken yazılı sınavlara göre kendimi daha rahat hissediyorum.					
36. Konuşma sınavı olmasaydı, derslerde konuşma etkinliklerine katılmazdım.					
37. Konuşma sınavı İngilizce düzeyimi geliştirmek için yararlıdır.					
38. Konuşma sınavı için çalıştığım birçok şeyi derslerde de kullanabilirim.					
39. Konuşma sınavı olmasaydı, konuşma becerimi iyileştirmek için bu kadar zaman harcamazdım.					
40. Konuşma sınavlarına daha sık girmem gerektiğini düşünüyorum.					
41. Konuşma sınavı beni geriyor.					
42. Konuşma sınavı en zor sınavdır.					
43. Konuşma sınavı kendimi en başarısız bulduğum sınavdır.					
44. Konuşma becerisini ölçmek zaman kaybı.					

## BÖLÜM 2

Aşağıdaki soruları dikkatlice okuyarak size uygun seçeneği işaretleyiniz ve boşlukları istenen bilgi ile doldurunuz.

1) Daha önce hiç konuşma sınavına girdiniz mi?

- a) Evet (ise yerini yazın)..... b) Hayır

2) Konuşma sınavına hazırlanmak için neler yaptınız? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |  |
|---|--|
| a) Dersi dikkatlice takip ettim.                                  | e) Günlük hayatımda konuşma pratikleri yaptım. |
| b) Sınava kendim, bireysel çalışmalarımı hazırladım.              | f) Hepsi                                       |
| c) Sınava arkadaşlarımla hazırladım.                              | g) Hiçbiri                                     |
| d) Profesyonel yardım (özel kurs veya İngilizce öğretmeni) aldım. | h) Diğer (.....)                               |

3) Konuşma sınavında kaç öğretmen görevliydi?

- a) 1      b) 2      c) 3      d) 3'ten fazla

4) Okulunuzda sınav ne zaman uygulanır? Birden fazla seçeneği işaretleyebilirsiniz.

- |                                |                               |
|--------------------------------|-------------------------------|
| a) Herhangi bir ders esnasında | d) İngilizce dersi esnasında  |
| b) Boş zamanımda               | e) Öğretmenimin boş zamanında |
| c) Okuldan sonra               | f) Diğer (.....)              |

5) Konuşma sınavı esnasında öğretmenleriniz hiç materyal kullandı mı?

- a) Evet (ise materyali yazınız) ..... b) Hayır

6) Konuşma sınavında hangi etkinlikleri yaptınız? Birden fazla seçeneği işaretleyebilirsiniz.

- a) Önceden verilen bir konu üzerine sunum yapma  
 b) Verilen resim hakkında konuşma  
 c) Kendini ve aileni tanıtmaya  
 d) Soru torbasından rastgele soru çekerek cevaplandırma  
 e) Çalışma amaçlı sınav öncesi verilen sorulara cevap verebilme  
 f) Bir diğer sınıf arkadaşıyla diyalog canlandırabilme  
 g) Sınıf arkadaşlarından oluşan bir grupta verilen konuyu tartışma  
 h) Rol yapma  
 i) İzlediği bir film / video üzerine konuşabilme (karakter, konu, yer, vb.)  
 j) Dinlediği veya okuduğu bir şey üzerine konuşabilme  
 k) Diğer (.....)



7) Konuşma sınavında hangi zorlukları yaşadınız? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |  |
|---|--|
| a) Soruları / etkinlikleri anlayabilme                      | g) Uygun kelimeyi doğru yerde kullanabilme   |
| b) Sorulara uygun cevap verebilme                           | h) Dil bilgisi kurallarına uygun konuşabilme |
| c) İlk Türkçe düşünüp sonra İngilizceye çevirerek cevaplama | i) Stresle baş edebilme                      |
| d) Akıcı konuşabilme  | j) Hepsi                                     |
| e) Konuşurken uygun kelimeleri hatırlama                    | k) Hiçbiri                                   |
| f) Sözcükleri doğru telaffuz edebilme                       | l) Diğer (.....)                             |

8) Size göre konuşma sınavında hangi ölçütler dikkate alınarak değerlendirme yapılmalıdır? Birden fazla seçeneği işaretleyebilirsiniz.

- |   |                  |
|---|------------------|
| a) İstenen konunun içeriğine göre konuşabilme       | g) Dil bilgisi   |
| b) Konuşmanın organizasyonu (içeriğin düzenlenmesi) | h) Vücut dili    |
| c) Konuşmanın akıcılığı                             | i) Göz teması    |
| d) Anlamsal netlike                                 | j) Kendine güven |
| e) Kelime bilgisi                                   | k) Hepsi         |
| f) Doğru telaffuz                                   | l) Hiçbiri       |
|   | m) Diğer (.....) |

**APPENDIX I: Teacher Interview Questions (Turkish Version)****GÖRÜŞME SORULARI**

- 1) Okulunuzda uygulanan konuşma sınavının olumlu bulduğunuz yönleri nelerdir?
- 2) Okulunuzda uygulanan konuşma sınavının olumsuz bulduğunuz yönleri nelerdir?
- 3) Okulunuzda uygulanan konuşma sınavını daha iyi hale getirmek için önerileriniz nelerdir?
- 4) Sizce iyi bir konuşma sınavı nasıl olmalıdır?
- 5) Güvenirliği artırmak için hangi önlemleri alırsınız?
  - a) Öğretmenler arası değerlendirme güvenilirliği ve aynı öğretmenin her öğrenci için değerlendirme güvenilirliğini nasıl sağlarsınız?
  - b) Sınavda aynı öğrenciye farklı notlar verirseniz, durumu nasıl ele alırsınız?

**APPENDIX J: Student Interview Questions (Turkish Version)****GÖRÜŞME SORULARI**

1. Okulunuzda uygulanan konuşma sınavının olumlu bulduğunuz yönleri nelerdir?
2. Okulunuzda uygulanan konuşma sınavının olumsuz bulduğunuz yönleri nelerdir?
3. Okulunuzda uygulanan konuşma sınavında aldığınız nottan memnun musunuz?
  - a. Memnun olmadığınızda öğretmeninizle görüşür müsünüz?
  - b. Öğretmeniniz bu durumu nasıl ele alır?
4. Okulunuzda uygulanan konuşma sınavıyla ilgili neyi değiştirmek istersiniz?
5. Okulunuzda uygulanan konuşma sınavında başka neler yapılabilir? Ne önerirsiniz?

## APPENDIX K: Permission from Denizli Directorate of National Education for the Questionnaire



T.C.  
DENİZLİ VALİLİĞİ  
İl Millî Eğitim Müdürlüğü

Sayı : 16605029/44-E.459005  
Konu : Anket İzni

12/01/2017

### VALİLİK MAKAMINA

İlgi : Pamukkale Üniversitesi Rektörlüğünün 02/01/2017 tarih ve 10 sayılı yazıları.

Üniversiteniz Eğitim Bilimleri Enstitüsü, Yabancı Diller Eğitimi Anabilim Dalı öğrencisi Ceyda ÖZDEMİR "Anadolu Liselerindeki 9. Sınıf Öğrencilerinin Konuşma Becerisinin Değerlendirilmesi Üzerine Bir Araştırma" konulu tezine yönelik hazırlanmış olduğu anket/ölçek formlarını İlgi yazı gereği Müdürlüğümüze bağlı Pamukkale ve Merkezefendi ilçelerinde bulunan anadolu liselerinde uygulamak istemektedir.

Yukarıda adı geçen müracaatlar ile ilgili (Lisans/Lisansüstü/Doktora) öğrencileri ve Öğretim Görevlilerinin ilgi yazıları ekinde belirtmiş oldukları okullarda, (Ortaöğretim/İlköğretim/Okulöncesi) konuları ile ilgili anket çalışmalarının "Araştırma, Yarışma ve Sosyal Etkinlik İzinleri" Genelgesinde belirtilen esaslar gereğince; Okul ve kurumların eğitim-öğretim faaliyetlerini aksatmayacak şekilde 2016/2017 eğitim-öğretim yılı içerisinde uygulamaları Müdürlüğümüze uygun görülmüştür.

Olurlarınıza arz ederim.

Hüseyin BAŞGÜN  
Millî Eğitim Müdür V.

OLUR  
12/01/2017  
Celalettin CANTÜRK  
Vali a.  
Vali Yardımcısı

T.C.  
DENİZLİ VALİLİĞİ  
İl Millî Eğitim Müdürlüğü

### PAMUKKALE ÜNİVERSİTESİ REKTÖRLÜĞÜNE

Kurumunuzca Müdürlüğümüzden talep edilen araştırma isteklerine ait Makam Onayı ve Müdürlüğümüze Onay verilen anket formları ekte gönderilmiştir.  
Gereğini rica ederim.

Celalettin CANTÜRK  
Vali a.  
Vali Yardımcısı

Ek:  
1-Anket Formları

Güvenli Elektronik İmza  
Aslı ile Aynıdır  
12.01.2017  
Mahmut TUR  
Memur

Şırakapılar Mh. Saltak Cd.No :76 Merkezefendi/DENİZLİ  
Elektronik Ağ: <http://denizli.meb.gov.tr>  
e-posta: [strateji20@meb.gov.tr](mailto:strateji20@meb.gov.tr)

Ayrıntılı bilgi için: R.ÇAN-Memur / S. GELMİŞ-ŞEF  
Tel: (0 258) 265 55 54 / 109-106  
Faks: (0 258) 265 01 69

**APPENDIX L: CV**

<b>Personal Information</b>	
Name	Ceyda
Surname	ÖZDEMİR
Birth date and place	02 November 1989 / Denizli
Nationality	Turkish
Address and e-mail adress	ceyda43662@hotmail.com
<b>Educational Background</b>	
Primary school	Osman Manisalı Primary School
High school	Lütfi Ege Anatolian Teacher Training High School
University	Dokuz Eylül University
<b>Professional Background</b>	
2011 – 2014	Van Erçek Mehmetçik Secondary School
2015 – Ongoing	Acıpayam Vocational and Technical Anatolian High School